

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/36586 A2

- | | |
|---|--|
| <p>(51) International Patent Classification⁷: C12N</p> <p>(21) International Application Number: PCT/IL00/00758</p> <p>(22) International Filing Date:
15 November 2000 (15.11.2000)</p> <p>(25) Filing Language: English</p> <p>(26) Publication Language: English</p> <p>(30) Priority Data:
09/443,338 19 November 1999 (19.11.1999) US</p> <p>(71) Applicant (<i>for all designated States except US</i>): YISSUM RESEARCH DEVELOPMENT COMPANY OF THE HEBREW UNIVERSITY OF JERUSALEM [IL/IL]; P.O. Box 4279, 46 Jabotinsky Street, 91042 Jerusalem (IL).</p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants (<i>for US only</i>): SHU, Wei [CN/IL]; 88/5 Etzel Road, 76361 Rehovot (IL). MARTON, Ira [IL/IL]; 10 Aharonovich Yoseph Street, 76347 Rehovot (IL). SIEGEL, Daniel, L. [IL/IL]; 11/5 Weizmann Street, 76280 Rehovot (IL). BEN-AMI, Bravdo [IL/IL]; 11 Han-kin Street, 76354 Rehovot (IL). DEKEL, Mara [IL/IL];</p> | <p>3 Lipkin Street, 76411 Rehovot (IL). SHOSEYOV, Oded [IL/IL]; 5 Haerez Street, 72910 Karme Yossef (IL).</p> <p>(74) Agent: G. E. EHRlich (1995) LTD.; 28 Bezalel Street, 52521 Ramat Gan (IL).</p> <p>(81) Designated States (<i>national</i>): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.</p> <p>(84) Designated States (<i>regional</i>): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Published:
— <i>Without international search report and to be republished upon receipt of that report.</i></p> <p><i>For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.</i></p> |
|---|--|



WO 01/36586 A2

(54) Title: **ASPEGILLUS NIGER BETA-GLUCOSIDASE GENE, PROTEIN AND USES THEREOF**

(57) Abstract: A polypeptide having β -glucosidase enzymatic activity, a polynucleotide encoding the polypeptide, a nucleic acid constructs carrying the polynucleotide, transformed or infected cells, such as yeast cells, and transgenic organisms expressing the polynucleotide and various uses of the polypeptide, the polynucleotide, cells and/or organisms, including, producing a recombinant polypeptide having the β -glucosidase enzymatic activity, increasing the level of aroma compounds in alcoholic beverages, as well as other fermentation products of plant material, hydrolyzing cellobiose and thus increasing the level of fermentable glucose, increasing the production of alcohol, such as ethanol from plant material, increasing the aroma released from a plant or a plant product, and hydrolysis or transglycosylation of glycosides.

ASPEGILLUS NIGER β -GLUCOSIDASE GENE, PROTEIN AND USES THEREOF

FIELD AND BACKGROUND OF THE INVENTION

5 The present invention relates to a polypeptide having β -glucosidase enzymatic activity, to a polynucleotide encoding the polypeptide, to nucleic acid constructs carrying the polynucleotide, to transformed or infected cells, such as yeast cells, and organisms expressing the polynucleotide and to various uses of the polypeptide, the polynucleotide, cells and/or organisms, including,
10 but not limited to, producing a recombinant polypeptide having β -glucosidase enzymatic activity, increasing the level of aroma compounds in alcoholic beverages, as well as other fermentation products of plant material, hydrolyzing cellobiose and thus increasing the level of fermentable glucose, to increase production of alcohol, such as ethanol from plant material, increasing
15 the aroma released from a plant or a plant product, and hydrolysis or transglycosylation of glycosides.

Abbreviations used herein include: BGL1 - *Aspergillus niger* B1 β -glucosidase; bgl1 - a cDNA encoding same; 2FGlcF - 2-deoxy-2-fluoro β -glucosyl fluoride; DNP - 2,4-dinitrophenol; DNPGlc - 2,4-dinitrophenyl β -D-glucopyranoside; pNP - p-nitrophenol; pNPGlc - p-nitrophenyl β -D-glucopyranoside; MUGlc - 4-methylumbelliferyl- β -D-glucopyranoside; YNB - yeast nitrogen base without amino acids; and X-glu - 5-bromo-4-chloro-3-indolyl β -D-glucopyranoside.
20

β -Glucosidases (EC 3.2.1.21; β -D-glucoside glucohydrolase) play a
25 number of different important roles in biology, including the degradation of cellulosic biomass by fungi and bacteria, degradation of glycolipids in mammalian lysosomes and the cleavage of glucosylated flavonoids in plants. These enzymes are therefore of considerable industrial interest, not only as constituents of cellulose-degrading systems, but also in the food industry (2,
30 3).

Aspergillus species are known as a useful source of β -glucosidases (4-6), and *Aspergillus niger* is by far the most efficient producer of β -glucosidase among the microorganisms investigated (4). Shoseyov *et al.* (7) have previously described a β -glucosidase from *Aspergillus niger* B1 (CMI CC 5 324626) which is active at low pHs, as well as in the presence of high ethanol concentrations. This enzyme effectively hydrolyzes flavor-compound glycosides in certain low-pH products, such as wine and passion fruit juice, thereby enhancing their flavor (8-12), and is particularly attractive for use in the food industry, as *A. niger* is considered non-toxic (3). In addition, β -glucosidase was found useful in enzymatic synthesis of glycosides (13-15). 10 Other *A. niger* β -glucosidases have also been purified (16-18), however, differences in their properties have been reported, including ranges of molecular weights (116-137 kDa), isoelectric points (pI values of 3.8-4) and pH optima (3.4-4.5). Indeed, at least two β -glucosidases, with distinct 15 substrate specificities, have been identified in commercial *A. niger* β -glucosidase preparations (19). Attempts to clear this confusion by cloning and expression of a functional *A. niger* β -glucosidase gene in *S. cerevisiae* has been previously reported (20), however the protein was not characterized, and the sequence was not published.

20 Glycosidases have been assigned to families on the basis of sequence similarities, there now being some 77 different such families defined containing over 2,000 different enzymes (21, see also <http://afmb.cnrs-mrs.fr/~pedro/CAZY/db.html>). With the exception of the glucosylceramidases (Family 30), all simple β -glucosidases belong to either Family 1 or 3. Family 1 contains enzymes from bacteria, plants and mammals, including also 6-phospho-glucosidases and thioglucosidases. Furthermore, most Family 1 25 enzymes also have significant galactosidase activity. Family 3 contains β -glucosidases and hexosaminidases of fungal, bacterial and plant origin. Enzymes of both families hydrolyze their substrates with net retention of

anomeric configuration, presumably via a two-step, double-displacement mechanism, involving two key active site carboxylic acid residues (for reviews of mechanism, see 22-24). In the first step, one of the carboxylic acids (the nucleophile) attacks at the substrate anomeric center, while the other (the acid/base catalyst) protonates the glycosidic oxygen, thereby assisting the departure of the aglycone. This results in the formation of a covalent α -glycosyl-enzyme intermediate. In a second step this intermediate is hydrolyzed by general base-catalyzed attack of water at the anomeric center of the glycosyl-enzyme, to release the β -glucose product and regenerate free enzyme. Both the formation and the hydrolysis of this intermediate proceed via transition states with substantial oxocarbenium ion character.

Given that Family 3 contains fungal enzymes of similar mass, including those from other *Aspergillus* sp., it is likely that the *Aspergillus niger* β -glucosidase would be a member of this family. Mechanistic information on this family is relatively sparse: the best characterized being the glycosylated 170 kDa β -glucosidase from *Aspergillus wentii*. By labeling the active site with conduritol B-epoxide, this enzyme was shown to carry out hydrolysis, with net retention of anomeric configuration. This study has demonstrated that the labeled aspartic acid residue was the same as that derivatized by the slow substrate D-glucal (1, 25). Furthermore, it was shown that the 2-deoxyglucosyl-enzyme, trapped by use of D-glucal, was kinetically identical to that formed during the hydrolysis of PNP-2-deoxy- β -D-glucopyranoside (26). Further detailed kinetic analysis of the enzyme was performed by Legler *et al.* (27), including measurement of Hammett relationships, kinetic isotope effects and studies of the binding of potent reversible inhibitors, such as gluconolactone and nojirimycin.

While reducing the present invention to practice, the β -glucosidase protein was isolated from *Aspergillus niger*, purified, cloned, sequenced, expressed in yeast host cells and its enzymatic function characterized. In

addition, the protein as well as signal peptide fused thereto and optionally an endoplasmic reticulum retaining peptide fused thereto were expressed in transgenic plants and the release of aroma substances therefrom following homogenization monitored. The enzyme encoded by the isolated gene, as described above, is of known usefulness in plant and/or plant products, as well as in biotechnological processes, including the food industry. Several unexpected advantages were uncovered, including, but not limited to, pH and temperature stability of the β -glucosidase from *Aspergillus niger*, requirement for a signal peptide for obtaining catalytic activity when expressed in plants. Advantage for an endoplasmic retaining peptide or for a lack thereof when expressed in plants, depending on the application.

SUMMARY OF THE INVENTION

According to one aspect of the present invention there is provided an isolated nucleic acid comprising a genomic, complementary or composite polynucleotide preferably being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an endoplasmic reticulum retaining peptide.

According to another aspect of the present invention there is provided a recombinant protein comprising a polypeptide having a β -glucosidase catalytic activity, the polypeptide is preferably derived from *Aspergillus niger* and it preferably fused to a signal peptide and optionally also to an endoplasmic reticulum retaining peptide.

According to yet another aspect of the present invention there is provided a nucleic acid construct comprising the isolated nucleic acid described herein.

According to still another aspect of the present invention there is provided host cell or an organism, such as a plant, comprising the nucleic acid or nucleic acid construct described herein.

According to further features in preferred embodiments of the invention described below, the polynucleotide is as set forth in SEQ ID NOs:1, 3 or a portion thereof.

According to still further features in the described preferred
5 embodiments, the nucleic acid construct further comprising at least one cis acting control element for regulating expression of the polynucleotide.

According to still further features in the described preferred embodiments, the host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.

10 According to still further features in the described preferred embodiments the prokaryotic cell is *E. coli*.

According to still further features in the described preferred embodiments the eukaryotic cell is selected from the group consisting of a yeast cell, a fungous cell, a plant cell and an animal cell.

15 According to still further features in the described preferred embodiments the polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having the β -glucosidase catalytic activity.

According to an additional aspect of the present invention there is provided a method of producing recombinant β -glucosidase, the method
20 comprising the step of introducing, in an expressible form, a nucleic acid construct into a host cell, the nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an
25 endoplasmic reticulum retaining peptide.

According to further features in preferred embodiments of the invention described below, the method further comprising the step of extracting the polypeptide having the β -glucosidase catalytic activity.

According to yet an additional aspect of the present invention there is provided a method of producing a recombinant β -glucosidase overexpressing cell, the method comprising the step of introducing, in an overexpressible form, a nucleic acid construct into a host cell, the nucleic acid construct
5 including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an endoplasmic reticulum retaining peptide.

According to still an additional aspect of the present invention there is
10 provided a method of increasing a level of at least one fermentation substance in a fermentation product, the method comprising the step of fermenting a glucose containing fermentation starting material by a yeast cell overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide being preferably derived from *Aspergillus niger*,
15 encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an endoplasmic reticulum retaining peptide, thereby increasing the level of the at least one fermentation substance in the fermentation product.

According to a further aspect of the present invention there is provided
20 a method of increasing a level of at least one fermentation substance in a fermentation product, the method comprising the step of fermenting a plant derived glucose containing fermentation starting material by a yeast cell, the plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from
25 *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an endoplasmic reticulum retaining peptide, thereby increasing the level of the at least one fermentation substance in the fermentation product.

According to a further aspect of the present invention there is provided
30 a method of increasing a level of at least one aroma substance in a plant

derived product, the method comprising the step of incubating a glucose containing plant starting material with a yeast cell overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an endoplasmic reticulum retaining peptide, thereby increasing the level of the at least one aroma substance in the plant derived product.

According to yet a further aspect of the present invention there is provided a method of increasing a level of at least one aroma substance in a plant derived product, the method comprising the step of incubating a glucose containing plant starting material with a yeast cell, said plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an endoplasmic reticulum retaining peptide, thereby increasing the level of the at least one aroma substance in the plant derived product.

According to still further features in the described preferred embodiments the plant derived product is a fermentation product, such as, but not limited to, an alcoholic beverage.

According to still a further aspect of the present invention there is provided a method of increasing a level of free glucose in a glucose containing fermentation starting material, the method comprising the step of fermenting the glucose containing fermentation starting material by a cell overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an endoplasmic reticulum retaining

peptide, thereby increasing the level of the free glucose in the glucose containing fermentation starting material.

According to another aspect of the present invention there is provided a method of increasing a level of free glucose in a plant derived glucose containing fermentation starting material, the method comprising the step of fermenting the plant derived glucose containing fermentation starting material by a cell, the plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an endoplasmic reticulum retaining peptide, thereby increasing the level of the free glucose in the plant.

According to yet another aspect of the present invention there is provided a method of increasing a level of free glucose in a plant, the method comprising the step of overexpressing in the plant a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an endoplasmic reticulum retaining peptide, thereby increasing the level of the free glucose in the plant.

According to still another aspect of the present invention there is provided a method of producing an alcohol, the method comprising the step of fermenting a glucose containing fermentation starting material by a cell overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an endoplasmic reticulum retaining peptide, and extracting the alcohol therefrom.

According to an additional aspect of the present invention there is provided a method of producing an alcohol, the method comprising the step of

fermenting a plant derived glucose containing fermentation starting material by a cell, the plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an endoplasmic reticulum retaining peptide, and extracting the alcohol therefrom.

According to an additional aspect of the present invention there is provided a method of producing an aroma spreading plant, the method comprising the step of overexpressing in the plant a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity and preferably further encoding, in frame, a signal peptide and an endoplasmic reticulum retaining peptide, thereby increasing aroma spread from the plant.

According to further features in preferred embodiments of the invention described below, overexpressing the nucleic acid construct is performed in a tissue specific manner.

According to still further features in the described preferred embodiments overexpressing the nucleic acid construct is limited to at least one tissue selected from the group consisting of flower, fruit, seed, root, stem, pollen and leaves.

The present invention successfully addresses the shortcomings of the presently known configurations by providing a polypeptide having β -glucosidase enzymatic activity, a polynucleotide encoding the polypeptide, a nucleic acid constructs carrying the polynucleotide, transformed or infected cells, such as yeast cells, and organisms expressing the polynucleotide and various uses of the polypeptide, the polynucleotide, cells and/or organisms, including, but not limited to, producing a recombinant polypeptide having β -glucosidase enzymatic activity, increasing the level of aroma compounds in alcoholic beverages, as well as other fermentation products of plant material,

hydrolyzing cellobiose and thus increasing the level of fermentable and/or free glucose, to increase production of a fermentation product, such as ethanol from plant material, increasing the aroma released from a plant or a plant product, and hydrolysis or transglycosylation of glycosides.

5

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is herein described, by way of example only, with reference to the accompanying drawings. With specific reference now to the drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of the preferred embodiments of the present invention only, and are presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of the invention. In this regard, no attempt is made to show structural details of the invention in more detail than is necessary for a fundamental understanding of the invention, the description taken with the drawings making apparent to those skilled in the art how the several forms of the invention may be embodied in practice.

In the drawings:

FIGs. 1a-c demonstrate plasmid maps employed as expression vectors for bgl1 cDNA. Figure 1a - *E. coli* expression vector containing bgl1 cDNA, inserted into the *NcoI/BamHI* sites of pET3d. Figure 1b - *S. cerevisiae* expression vector containing bgl1 cDNA, inserted into the *HindIII/BamHI* sites of pYES2-bgl1 plasmid. Figure 1c - *P. pastoris* expression vector containing bgl1 cDNA, inserted into the *EcoRI/BamHI* sites of pHIL-S1.

FIGs. 2a-b demonstrates SDS-PAGE analysis of active protein samples eluted from a Mono-Q column, stained with coomassie blue (Figure 2a), or β -glucosidase zymogram (Figure 2b) using MUGlc as a substrate. Lanes (for both Figures 2a and 2b): 1 - Electroeluted band of BGL1 from preparative PAGE-SDS gel stabs; 2, 3, 4, 5 - acetone precipitates from Mono-Q separation of BGL1.

FIG. 3 demonstrates SDS-PAGE analysis of purified β -glucosidase by Mono-Q and Resource-S. Lanes: 1 - crude (27.5 μ g protein); 2 - active fraction after Mono-Q (7 μ g protein); and 3 - active fraction after Resource-S (10 μ g protein).

5 FIG. 4 demonstrates SDS-PAGE analysis of β -glucosidase deglycosylated by N-glycosidase-F. Lanes: 1 - molecular weight marker; 2 - native β -glucosidase; and 3 - deglycosylated protein.

FIG. 5a demonstrates the DNA and amino acid sequences of *bgl1*. Amino acid sequences determined by Edman degradation are underlined. DNA sequences of introns are underlined. Signal peptide is indicated by italic letters.

FIG 5b. demonstrates *bgl1* gene organization. Exons (E1-7) are indicated by filled boxes, introns by solid lines, restriction sites and the stop codon by arrows.

15 FIG. 6a demonstrates a Western blot analysis of recombinant BGL1 expressed in *S. cerevisiae*. Lanes: 1 - native BGL1 (positive control); 2 - total protein extract of *S. cerevisiae* expressing recombinant BGL1; 3 - total protein extract of *S. cerevisiae* without the *bgl1* expression vector (negative control).

FIG. 6b demonstrates a Western blot analysis of recombinant BGL1 secreted from *P. pastoris*. Lanes: 1 - molecular weight marker; 2 - medium supernatant of *P. pastoris* expressing recombinant BGL1; 3 - medium supernatant of *P. pastoris* host without the vector (negative control).

FIG. 7 demonstrates proton-NMR spectra, illustrating the stereochemical course of pNPGlc hydrolysis by *A. niger* β -glucosidase. Spectra are for the anomeric proton region of the substrate at different time intervals relative to addition of the enzyme.

FIG. 8 demonstrates inactivation of recombinant BGL1 by 2FGlcF. Pure enzyme was incubated in the presence of various concentrations of the inactivator, and residual enzyme activity was determined at different time

intervals. Residual activity is presented, semilogarithmically, versus time, in the presence of the indicated concentrations of inactivator.

FIG. 9 demonstrates reactivation of 2-deoxy-2-fluoroglucosyl-recombinant BGL1 by linamarin. Activity is plotted versus incubation time in the presence of the indicated concentrations of linamarin.

FIG. 10 demonstrates the stability of recombinant *A. niger* β -glucosidase at various temperatures. Activity is calculated as percent of a recombinant enzyme solution kept at 4 °C.

FIGs. 11a-c show schematic depictions of expression cassettes used for expression of *A. niger* β -glucosidase in tobacco plants. Figure 11a - a cassette encoding BGL1 without a signal peptide (see, SEQ ID NO:13 for the nucleotide sequence and SEQ ID NO:14 for the amino acid sequence); Figure 11b - a cassette encoding a BGL1 fused to a Cell signal peptide for secretion into the apoplast (see, SEQ ID NO:15 for the nucleotide sequence and SEQ ID NO:16 for the amino acid sequence); and Figure 11c - a cassette encoding a BGL1 fused to Cell signal peptide as in Figure 11b and in addition to HDEL (SEQ ID NO:17) ER-retaining peptide at the C-terminus for accumulation in the ER (see, SEQ ID NO:18 for the nucleotide sequence and SEQ ID NO:19 for the amino acid sequence).

FIG. 12 demonstrate PCR amplification results of *bgl1* cDNA indicating the presence of *bgl1* cDNA in transgenic plants. CB10 and CB11 - transgenic plants transformed with *bgl1* and Cell signal peptide without HDEL ER retaining peptide. CBT3, CBT8 and CBT15 - different transgenic lines transformed with *bgl1*, Cell signal peptide and HDEL. B1 - a transgenic plants transformed with *bgl1*. 1kb - 1 kb DNA marker. WT - wild type non transgenic plant. pETB1 - *bgl1* plasmid DNA.

FIGs. 13a-b show Western blot analyses of transgenic plants containing BGL1 without signal peptide (13a), and BGL1 with Cell signal peptide (13b), with and without HDEL ER retaining peptide. An gluco - purified *A. niger* beta-glucosidase. WT - nontransgenic control plant. B1, B15, B16, B20, B27,

B33 and B34 - different transgenic lines transformed with bgl1. CBT1, CBT 3, CBT 7 and CBT 8 - different transgenic lines transformed with bgl1, Cell signal peptide and HDEL. CB10 and CB12 - transgenic plants transformed with bgl1 and Cell signal peptide without HDEL ER retaining peptide.

FIG. 14 show activity gel analysis of transgenic tobacco plant extracts in SDS-PAGE incubated with MUGlu. WT - non-transgenic control plant. CB10 and CB11 - two independent lines of transgenic plants expressing BGL1 fused to Cell signal peptide (without HDEL). CBT3, CBT8 and CBT15 - independent lines of transgenic plants expressing BGL1 fused to Cell signal peptide at the N terminus and HDEL ER retaining peptide at the C terminus. B1 and B34 - transgenic plant expressing BGL1 without signal peptide or HDEL ER retaining peptide and which were positive for BGL1 protein in Western blot analysis. An Glu - control *A. niger* native beta-glucosidase.

FIG. 15 demonstrates level of BGL1 activity in different transgenic plants. WT - non-transgenic control plant. B1 and B21 - transgenic plants expressing BGL1 without signal peptide or HDEL ER retaining peptide and which were positive for BGL1 in Western blot analysis. CBT8, CBT21, CBT0 and CBT15 - independent lines of transgenic plants expressing BGL1 fused to Cell signal peptide at the N terminus and HDEL ER retaining peptide at the C terminus. CB12, CB13, CB14 and CB15 - four independent lines of transgenic plants expressing BGL1 fused to Cell signal peptide (without HDEL).

DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention is of a polypeptide having β -glucosidase enzymatic activity, a polynucleotide encoding the polypeptide, a nucleic acid constructs carrying the polynucleotide, transformed or infected cells, such as yeast cells, and organisms expressing the polynucleotide and various uses of the polypeptide, the polynucleotide, cells and/or organisms, including, but not limited to, producing a recombinant polypeptide having the β -glucosidase

enzymatic activity, increasing the level of aroma compounds in alcoholic beverages, as well as other fermentation products of plant material, hydrolyzing cellobiose and thus increasing the level of fermentable glucose, increasing the production of alcohol, such as ethanol from plant material, increasing the aroma released from a plant or a plant product, and hydrolysis or transglycosylation of glycosides.

The principles and operation of the present invention may be better understood with reference to the drawings and accompanying descriptions.

Before explaining at least one embodiment of the invention in detail, it is to be understood that the invention is not limited in its application to the details of the components set forth in the following description or exemplified in the examples that follow. The invention is capable of other embodiments or of being practiced or carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein is for the purpose of description and should not be regarded as limiting.

According to one aspect of the present invention there is provided an isolated nucleic acid comprising a genomic, complementary or composite polynucleotide encoding a polypeptide having a β -glucosidase catalytic activity. Preferably the polynucleotide is derived from *Aspergillus niger*, however other sources are applicable. These include all isolated polynucleotides encoding polypeptide having β -glucosidase catalytic activity. Such polynucleotides and polypeptides identified by their GenBank Accession Nos. are listed in Table 1 below, all of which can be used while implementing the present invention.

TABLE 1

Accession numbers of cDNA and their encoded beta-glucosidases (EC.3.2.1.21)

Organism	SWISS-PROT	EMBL
<i>Acetobacter xylinus</i>	O24749	AB003689; AB010645
<i>Agrobacterium sp.</i>	P12614	M19033; AAA22085.1
<i>Agrobacterium tumefaciens</i>	P27034	M59852; AAA22082.1

<i>Arabidopsis thaliana</i>	O82772, O24433, O23656	AF082157; AF082158; AC009327; U72153; U72155 AC020665; AC066691
<i>Aspergillus aculeatus</i>	P48825	D64088, BAA10968.1
<i>Aspergillus kawachi</i>	P87076	AB003470
<i>Aspergillus niger</i> B1		AJ132386; CAB75696.1
<i>Aspergillus niger</i> AMS1	Q9P456	AF268911
<i>Avena sativa</i>	Q38786, Q9ZP27	X78433; AF082991
<i>Azospirillum irakense</i>		AF090429; AAF21798.1
<i>Bacillus circulans</i>	Q03506	M96979; AAA22266.1
<i>Bacillus</i> sp. GL1	Q9ZNN7	AB009411; BAA36161.1; AB009410
<i>Bacillus polymyxa</i>	P22073, P22505	M60210; M60211
<i>Bacillus subtilis</i>	P40740	Z34526; CAA84287.1
<i>Bacillus subtilis</i>	P42403	D30762; BAA06429.1
<i>Bacteroides fragilis</i>	O31356	AF006658; AAB62870.1
<i>Bifidobacterium breve</i>	P94248, O08487	D84489; D88311
<i>Botryotinia fuckeliana</i>		AJ130890; CAB61489.1
<i>Brassica napus</i>	Q42618	X82577
<i>Brassica nigra</i>	O24434	U72154
<i>Butyrivibrio fibrisolvens</i>	P16084	M31120; AAA23008.1
<i>Caldocellum saccharolyticum</i>	P10482	X12575; CAA31087.1
<i>Caldicellulosiruptor</i> sp. 14B	Q9ZEN0	AJ131346
<i>Candida wickerhamii</i>	Q12601	U13672
<i>Cavia porcellus</i>	P97265	U50545
<i>Cellulomonas biazotea</i>	O51843	AF005277; AAC38196.1
<i>Cellulomonas fimi</i>	Q46043	M94865
<i>Cellvibrio gilvus</i>	P96316	D14068; BAA03152.1
<i>Chryseobacterium meningosepticum</i>	O30713	AF015915
<i>Clostridium stercorarium</i>	O08331	Z94045
<i>Clostridium thermocellum</i>	P26208	X60268; CAA42814.1
<i>Clostridium thermocellum</i>	P14002	X15644; CAA33665.1
<i>Coccidioides immitis</i>	O14424	U87805; AF022893
<i>Costus speciosus</i>	Q42707	D83177
<i>Dalbergia cochinchinensis</i>	Q9SPK3	AF163097
<i>Dictyostelium discoideum</i>	Q23892	L21014
<i>Digitalis lanata</i>	Q9ZPB6	AJ133406
<i>Erwinia chrysanthemi</i>	Q46684	U08606; AAA80156.1
<i>Erwinia herbicola</i>	Q59437	X79911; CAA56282.1
<i>Escherichia coli</i>	P33363	U15049; AAB38487.1
<i>Escherichia coli</i> K12/MG1655	E65074, Q46829	U28375; AE000373
<i>Glycine max</i>		AF000378; AAD09291.1
<i>Hansenula anomala</i>	P06835	X02903; CAA26662.1
<i>Homo sapiens</i>		AJ278964; CAC08178.1
<i>Hordeum vulgare</i>	Q40025	L41869
<i>Humicola grisea</i> var. <i>thermoidea</i>	O93784	AB003109
<i>Kluyveromyces marxianus</i>	P07337	X05918; CAA29353.1
<i>Lactobacillus plantarum</i>	O86291	Y15954; AJ250202; CAB71149.1
<i>Manihot esculenta</i>	Q40283	X94986
<i>Microbispora bispora</i>	P38645	M97265; AAA25311.1
<i>Nicotiana tabacum</i>	O82151	AB017502; BAA33065.1
<i>Orpinomyces</i> sp. PC-2		AF016864; AAD45834.1
<i>Oryza sativa</i>	Q42975	U28047

<i>Paenibacillus polymyxa</i>	P22073	M60210; AAA22263.1
<i>Paenibacillus polymyxa</i>	P22505	M60211; AAA22264.1
<i>Phaeosphaeria avenaria</i>		AJ276675; CAB82861.1
<i>Phanerochaete chrysosporium</i>	O74203	AF036872; AF036873
<i>Pichia anomala</i> (Candida pelliculosa)	P06835	X02903
<i>Pinus contorta</i>		AF072736; AAC696.1
<i>Polygonum tinctorium</i>		AB003089; BAA78708.1
<i>Prunus avium</i>	Q43014	U39228
<i>Prunus serotina</i>	Q43073, Q40984	U50201; U26025
<i>Prevotella albensis</i> M384		AJ276021; CAC07184.1
<i>Prevotella ruminicola</i>	Q59716	U35425
<i>Pyrococcus furiosus</i>	Q51723	AF013169; U37557
<i>Ruminococcus albus</i>	P15885 O66050	X15415; CAA33461.1 U92808
<i>Saccharomycopsis fibuligera</i>	P22506	M22475; AAA34314.1
<i>Saccharomycopsis fibuligera</i>	P22507	M22476; AAA34315.1
<i>Saccharopolyspora erythraea</i>	O70021	Y14327
<i>Salmonella typhimurium</i>	Q56078	D86507; BAA13102.1
<i>Schizophyllum commune</i>	P29091	M27313; AAA33925.1
<i>Schizosaccharomyces pombe</i>		AL355920; CAB91163.1
<i>Secale cereale</i>		AF293849; AAG00614.1
<i>Septoria lycopersici</i>	Q99324	U24701; U35462
<i>Sorghum bicolor</i>	Q41290	U33817
<i>Spodoptera frugiperda</i>	O61594	AF052729
<i>Streptomyces coelicolor</i> A3(2)		AL121596; CAB56653.1
<i>Streptomyces reticuli</i>	Q9X9R4	AJ009797
<i>Streptomyces rochei</i> A2	Q55000	X74291
<i>Streptomyces</i> sp. QM-B814	Q59976	Z29625
<i>Thermoanaerobacter brockii</i>	P96090, Q60026	Z56279; Z56279
<i>Thermobifida fusca</i> ER1		AF086819; AAF37727.1
<i>Thermococcus</i> sp.	O08324	Z70242
<i>Thermotoga maritima</i>	Q08638	X74163; CAA52276.1
<i>Thermotoga neapolitana</i>	Q33843, Q60038	Z97212; Z77856; CAB10165.1
<i>Thermus</i> sp. Z-1	Q9RA58	AB034947
<i>Thermus thermophilus</i>	Q9X9D4	Y16753
<i>Trichoderma reesei</i> (Hypocrea jecorina)	Q12715, O93785	U09580; AAA18473.1, AB003110
<i>Trifolium repens</i>	P26204	X56734; CAA40058.1
<i>Trifolium repens</i>	P26205	X56733; CAA40057.1
<i>Tropaeolum majus</i>	O82074	AJ006501; CAA07070.1
<i>Zea mays</i>	P49235, Q41761	X74217, U25157; CAA52293.1 U33816, U44087, U44773
Unidentified bacterium	Q60055	U12011

As used herein in the specification and in the claims section that follows, the term "isolated" refers to a biological component (such as a nucleic acid or protein or organelle) that has been substantially separated or purified away from other biological components in the cell of the organism in which the component naturally occurs, i.e., other chromosomal and extra-

chromosomal DNA and RNA, proteins and organelles. Nucleic acids and proteins that have been "isolated" include nucleic acids and proteins purified by standard purification methods. The term also embraces nucleic acids and proteins prepared by recombinant expression in a host cell as well as
5 chemically synthesized nucleic acids.

As used herein and in the claims section that follows the terms and phrases "polynucleotide" and "polynucleotide sequence" are used interchangeably and refer to a nucleotide sequence which can be DNA or RNA of, for example, genomic or synthetic origin, which may be single- or double-
10 stranded, and which may represent the sense or antisense strand. Similarly, the terms "polypeptide" and "polypeptide sequence" are interchangeably used herein and refer to an amino acid sequence of any length.

As used herein in the specification and in the claims section that follows, the phrase "complementary polynucleotide sequence" includes
15 sequences, which originally result from reverse transcription of messenger RNA using a reverse transcriptase or any other RNA dependent DNA polymerase. Such sequences can be subsequently amplified *in vivo* or *in vitro* using a DNA dependent DNA polymerase.

As used herein in the specification and in the claims section that follows, the phrase "genomic polynucleotide sequence" includes sequences
20 which originally derive from a chromosome and reflect a contiguous portion of a chromosome.

As used herein in the specification and in the claims section that follows, the phrase "composite polynucleotide sequence" includes sequences
25 which are at least partially complementary and at least partially genomic. A composite sequence can include some exonal sequences required to encode the polypeptide having the β -glucosidase catalytic activity, as well as some intronic sequences interposing therebetween. The intronic sequences can be of any source, including of other genes, and typically will include conserved

splicing signal sequences. Such intronic sequences may further include *cis* acting expression regulatory elements, as hereinbelow described.

As used herein in the specification and in the claims section that follows, the phrase "having a β -glucosidase catalytic activity" refers to a polypeptide sequence, protein or fragments thereof capable of serving as catalysts to a chemical reaction involving hydrolysis of the O-glycosidic bond of glucosides, the result of which is the release of a β -D-glucose residue(s), or an aglycon, in addition to the β -D-glucose residue. Specifically, hydrolysis by retaining enzymes is performed while maintaining the β -configuration of the anomeric center of the carbohydrate. A wide specificity for β -glucosides exists, thus, some examples also hydrolyze one or more of the following: β -D-galactosides, α -L- arabinosides, β -D-xylosides, and β -D-fucosides.

As used herein the term "catalyst" refers to a substance that accelerates a chemical reaction, but is not consumed or changed permanently thereby.

As used herein the term "glucoside" refers to a compound of at least two monomers, at least one of which is a glucose, including a glycoside bond. Examples of glucosides include, but are not limited to, glucose containing backbones, such as the diglucose cellobiose, and the glucose polymer, cellulose.

According to preferred embodiments, the polynucleotide according to this aspect of the present invention encodes a polypeptide as set forth in SEQ ID NO:2 or a portion thereof which retains β -glucosidase catalytic activity.

Alternatively or additionally, the polynucleotide according to this aspect of the present invention is as set forth in SEQ ID NO: 1, 3 or a portion thereof, the portion encodes a polypeptide retaining β -glucosidase catalytic activity.

In a broader aspect the polynucleotides according to the present invention encode a polypeptide which is at least 75 %, at least 80 %, at least 85 %, at least 90 %, at least 95 % or more, say 95 % - 100 % homologous to SEQ ID NO:2 as determined using the BestFit software of the Wisconsin sequence

analysis package, utilizing the Smith and Waterman algorithm, where gap creation penalty equals 8 and gap extension penalty equals 2.

According to preferred embodiments, the polynucleotides according to the broader aspect of the present invention encodes a polypeptide as set forth
5 in SEQ ID NOs:1 or 3 or a portion thereof which retains activity.

Alternatively or additionally, the polynucleotides according to this broader aspect of the present invention are hybridizable with SEQ ID NOs: 1 or 3.

Hybridization for long nucleic acids (e.g., above 200 bp in length) is
10 effected according to preferred embodiments of the present invention by stringent or moderate hybridization, wherein stringent hybridization is effected by a hybridization solution containing 10 % dextrane sulfate, 1 M NaCl, 1 % SDS and 5×10^6 cpm ^{32}p labeled probe, at 65 °C, with a final wash solution of 0.2 x SSC and 0.1 % SDS and final wash at 65°C; whereas moderate
15 hybridization is effected by a hybridization solution containing 10 % dextrane sulfate, 1 M NaCl, 1 % SDS and 5×10^6 cpm ^{32}p labeled probe, at 65 °C, with a final wash solution of 1 x SSC and 0.1 % SDS and final wash at 50 °C.

Yet alternatively or additionally, the polynucleotides according to this broad aspect of the present invention is preferably at least 70 %, at least 75 %, at least 80 %, at least 85 %, at least 90 %, at least 95 % or more, say 95 % -
20 100 %, identical with SEQ ID NOs: 1 or 3 as determined using the BestFit software of the Wisconsin sequence analysis package, utilizing the Smith and Waterman algorithm, where gap weight equals 50, length weight equals 3, average match equals 10 and average mismatch equals -9.

25 Thus, this broad aspect of the present invention encompasses (i) polynucleotides as set forth in SEQ ID NOs:1 or 3; (ii) fragments thereof; (iii) sequences hybridizable therewith; (iv) sequences homologous thereto; (v) sequences encoding similar polypeptides with different codon usage; (vi) altered sequences characterized by mutations, such as deletion, insertion or

substitution of one or more nucleotides, either naturally occurring or man induced, either randomly or in a targeted fashion.

According to another aspect of the present invention there is provided a nucleic acid construct comprising the isolated nucleic acid described herein.

5 According to a preferred embodiment, the nucleic acid construct according to this aspect of the present invention further comprising at least one *cis* acting control (regulatory) element for regulating the expression of the isolated nucleic acid. Such *cis* acting regulatory elements include, for example, promoters, which are known to be sequence elements required for
10 transcription, as they serve to bind DNA dependent RNA polymerase, which transcribes sequences present downstream thereof. Further details relating to various regulatory elements are described hereinbelow.

While the isolated nucleic acid described herein is an essential element of the invention, it is modular and can be used in different contexts. The
15 promoter of choice that is used in conjunction with this invention is of secondary importance, and will comprise any suitable promoter. It will be appreciated by one skilled in the art, however, that it is necessary to make sure that the transcription start site(s) will be located upstream of an open reading frame. In a preferred embodiment of the present invention, the promoter that
20 is selected comprises an element that is active in the particular host cells of interest. These elements may be selected from transcriptional regulators that activate the transcription of genes essential for the survival of these cells in conditions of stress or starvation, including the heat shock proteins.

A construct according to the present invention preferably further
25 includes an appropriate selectable marker. In a more preferred embodiment according to the present invention the construct further includes an origin of replication. In another most preferred embodiment according to the present invention the construct is a shuttle vector, which can propagate both in *E. coli* (wherein the construct comprises an appropriate selectable marker and origin
30 of replication) and be compatible for propagation in cells, or integration in the

genome, of an organism of choice, such as a plant. The construct according to this aspect of the present invention can be, for example, a plasmid, a bacmid, a phagemid, a cosmid, a phage, a virus or an artificial chromosome.

According to an additional aspect of the present invention there is
5 provided a recombinant protein comprising a polypeptide having a β -glucosidase catalytic activity. The polypeptide is preferably derived from an *Aspergillus niger* and preferably includes a signal peptide and optionally an endoplasmic reticulum retaining peptide.

According to preferred embodiments, the polypeptide according to this
10 aspect of the present invention is as set forth in SEQ ID NO:2 or a portion thereof which retains β -glucosidase catalytic activity.

SEQ ID NO:2 of *A. niger* β -glucosidase is similar to the amino acid sequence of the β -glucosidase of *A. kawachii*. However, while the former is highly stable at wide range of temperatures and pH treatments, the latter is
15 relatively unstable, and thus has certain disadvantages, rendering its use for the purpose of the present invention as is further detailed and described hereinunder, unfeasible and/or much less attractive.

Recently, Iwashita and coworkers have published the sequence of a β -glucosidase (GenBank/EMBL AB003470) obtained from *Aspergillus kawachii*
20 strain: IFO4308. Sequence comparison between *Aspergillus kawachii* β -glucosidase and *A. niger* β -glucosidase revealed that the two share 98 % homology.

Enzymes of the two *Aspergillus* sp. contain seven cysteine residues and identical number of glycosylation sites, while differing in their degree of
25 glycosylation (35).

The physical and kinetic properties of three β -glucosidases from *Aspergillus kawachii* were described (35), and the three were shown to be products of the same gene, differing solely by the degree of glycosylation. The three purified *A. kawachii* β -glucosidases were readily inactivated, even at

moderate pH and temperature conditions. In sharp distinction, while examining the stability of the recombinant *A. niger* β -glucosidase according to the present invention under conditions identical to those described by Iwashita et al. and as described hereinbelow in the Examples section, revealed that the enzyme is highly stable, retaining majority of the enzymatic activity even after 1 hour incubation at 60 °C (68 % activity, as defined by percent activity of an enzyme kept at 4 °C).

Thus, despite the similarity between the *A. kawachii* and *A. niger* β -glucosidases, the *A. niger* enzyme unexpectedly exhibits significantly higher thermal and pH stability.

According to yet another aspect of the present invention there is provided a host cell comprising a nucleic acid construct as described herein. The term "host cell" refers to a recipient of a heterologous nucleic acid, which host cell can be either a prokaryotic cell, such as *E. coli*, or a eukaryotic cell, such as a yeast cell, a filamentous fungus cell, a plant cell or an animal cell. Examples for a yeast cell include, but not limited to, *Pichia* sp. such as *P. pastoris*, and *Saccharomyces* sp. such as *S. cerevisiae*.

As used herein and in the claims section which follows, the term "heterologous" when used in context of a nucleic acid sequence or a protein found within a plant, plant derived tissue or plant cells, or alternatively, within a eukaryotic cell, such as yeast, or a prokaryotic cell such as bacteria, refers to nucleic acid or amino acid sequences typically not native to the plant, plant derived tissue or plant cells, or alternatively, to the eukaryotic cell, such as yeast, or the prokaryotic cell, such as bacteria. Interchangeably, nucleic acid or amino acid sequences typically not native to the plant, plant derived tissue or plant cells, or alternatively, to the eukaryotic cell, such as yeast, or the prokaryotic cell, such as bacteria, are referred to by "recombinant nucleic acid" and "recombinant protein", respectively. Thus, a recombinant nucleic acid is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of

sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques.

As used herein in the specification and in the claims section that follows, the term "eukaryotic cell" refers to a cell containing a diploid genome through at least a portion of its life cycle, having membrane-bound nucleus with chromosomes made of DNA, with cell division involving a form of mitosis in which spindles are involved. Possession of a eukaryote type of cell characterizes the four kingdoms, *Protoctista*, *Fungi*, *Plantae* and *Animalia*.

As used herein in the specification and in the claims section that follows, the term "prokaryotic cell" refers to various bacteria and blue-green algae, characterized by the absence of the nuclear organization, mitotic capacities and complex organelles that typify the eukaryote superkingdom. Examples of prokaryotic cell according to the present invention are bacteria, such as, but not limited to, *E. coli*.

According to still another aspect of the present invention there is provided an organism comprising a nucleic acid construct as described herein, such as, but not limited to, a plant. Such an organism is said to be transformed or virally infected.

As used herein the term "transformed" and its conjugations such as transformation, transforming and transform, all relate to the process of introducing heterologous nucleic acid sequences into a cell or an organism, which nucleic acid sequences are propagatable to the offspring. The term thus reads on, for example, "genetically modified", "transgenic" and "transfected", which may be used herein to further describe and/or claim the present invention. The term relates both to introduction of a heterologous nucleic acid sequence into the genome of an organism and/or into the genome of a nucleic acid containing organelle thereof, such as into a genome of chloroplast or a mitochondrion.

As used herein the phrase "viral infected" includes infection by a virus carrying a heterologous nucleic acid sequence. Such infection typically results in transient expression of the nucleic acid sequence, which nucleic acid sequence is typically not integrated into a genome and therefore not propagatable to offspring, unless further infection of such offspring is experienced.

There are various methods of introducing foreign genes into both monocotyledonous and dicotyledonous plants (Potrykus, I., *Annu. Rev. Plant. Physiol., Plant. Mol. Biol.* (1991) 42:205-225; Shimamoto *et al.*, *Nature* (1989) 338:274-276). The principle methods of causing stable integration of exogenous DNA into plant genomic DNA include two main approaches:

(i) *Agrobacterium*-mediated gene transfer: Klee *et al.* (1987) *Annu. Rev. Plant Physiol.* 38:467-486; Klee and Rogers in *Cell Culture and Somatic Cell Genetics of Plants*, Vol. 6, *Molecular Biology of Plant Nuclear Genes*, eds. Schell, J., and Vasil, L. K., Academic Publishers, San Diego, Calif. (1989) p. 2-25; Gatenby, in *Plant Biotechnology*, eds. Kung, S. and Arntzen, C. J., Butterworth Publishers, Boston, Mass. (1989) p. 93-112.

(ii) direct DNA uptake: Paszkowski *et al.*, in *Cell Culture and Somatic Cell Genetics of Plants*, Vol. 6, *Molecular Biology of Plant Nuclear Genes* eds. Schell, J., and Vasil, L. K., Academic Publishers, San Diego, Calif. (1989) p. 52-68; including methods for direct uptake of DNA into protoplasts, Toriyama, K. *et al.* (1988) *Bio/Technology* 6:1072-1074. DNA uptake induced by brief electric shock of plant cells: Zhang *et al.* *Plant Cell Rep.* (1988) 7:379-384. Fromm *et al.* *Nature* (1986) 319:791-793. DNA injection into plant cells or tissues by particle bombardment, Klein *et al.* *Bio/Technology* (1988) 6:559-563; McCabe *et al.* *Bio/Technology* (1988) 6:923-926; Sanford, *Physiol. Plant.* (1990) 79:206-209; by the use of micropipette systems: Neuhaus *et al.*, *Theor. Appl. Genet.* (1987) 75:30-36; Neuhaus and Spangenberg, *Physiol. Plant.* (1990) 79:213-217; or by the direct incubation of DNA with germinating pollen, DeWet *et al.* in *Experimental*

Manipulation of Ovule Tissue, eds. Chapman, G. P. and Mantell, S. H. and Daniels, W. Longman, London, (1985) p. 197-209; and Ohta, Proc. Natl. Acad. Sci. USA (1986) 83:715-719.

The *Agrobacterium* system includes the use of plasmid vectors that contain defined DNA segments that integrate into the plant genomic DNA. Methods of inoculation of the plant tissue vary depending upon the plant species and the *Agrobacterium* delivery system. A widely used approach is the leaf disc procedure, which can be performed with any tissue explant that provides a good source for initiation of whole plant differentiation. Horsch *et al.* in Plant Molecular Biology Manual A5, Kluwer Academic Publishers, Dordrecht (1988) p. 1-9. A supplementary approach employs the *Agrobacterium* delivery system in combination with vacuum infiltration. The *Agrobacterium* system is especially viable in the creation of transgenic dicotyledenous plants.

There are various methods of direct DNA transfer into plant cells. In electroporation, the protoplasts are briefly exposed to a strong electric field. In microinjection, the DNA is mechanically injected directly into the cells using very small micropipettes. In microparticle bombardment, the DNA is adsorbed on microprojectiles such as magnesium sulfate crystals or tungsten particles, and the microprojectiles are physically accelerated into cells or plant tissues.

Following transformation plant propagation is exercised. The most common method of plant propagation is by seed. Regeneration by seed propagation, however, has the deficiency that due to heterozygosity there is a lack of uniformity in the crop, since seeds are produced by plants according to the genetic variances governed by Mendelian rules. Basically, each seed is genetically different and each will grow with its own specific traits. Therefore, it is preferred that the transformed plant be produced such that the regenerated plant has the identical traits and characteristics of the parent transgenic plant. Therefore, it is preferred that the transformed plant be

regenerated by micropropagation, which provides a rapid, consistent reproduction of the transformed plants.

Micropropagation is a process of growing new generation plants from a single piece of tissue that has been excised from a selected parent plant or
5 cultivar. This process permits the mass reproduction of plants having the preferred tissue expressing the protein. The new generation plants, which are produced, are genetically identical to, and have all of the characteristics of, the original plant. Micropropagation allows mass production of quality plant material in a short period of time and offers a rapid multiplication of selected
10 cultivars in the preservation of the characteristics of the original transgenic or transformed plant. The advantages of cloning plants are the speed of plant multiplication and the quality and uniformity of plants produced.

Micropropagation is a multi-stage procedure that requires alteration of culture medium or growth conditions between stages. Thus, the
15 micropropagation process involves four basic stages: Stage one, initial tissue culturing; stage two, tissue culture multiplication; stage three, differentiation and plant formation; and stage four, greenhouse culturing and hardening. During stage one, initial tissue culturing, the tissue culture is established and certified contaminant-free. During stage two, the initial tissue culture is
20 multiplied until a sufficient number of tissue samples are produced to meet production goals. During stage three, the tissue samples grown in stage two are divided and grown into individual plantlets. At stage four, the transformed plantlets are transferred to a greenhouse for hardening where the plants' tolerance to light is gradually increased so that it can be grown in the natural
25 environment.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression
30 cassette into a plant genome.

Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The constructs of the subject invention will include an expression
5 cassette for expression of the protein of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous sequence one or more of the following sequence elements, a promoter region, plant 5' untranslated sequences which can include regulatory elements, initiation codon
10 depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

As used herein, the phrase "regulatory element" refers to a nucleotide
15 sequence which are typically included within an expression cassette and function in regulating (i.e., enhancing or depressing) the expression of a coding sequence therefrom. This regulation can be effected either at the transcription or the translation stages. Examples of regulatory elements include, but are not limited to, enhancers, suppressers and transcription
20 terminators.

As used herein the term "promoter" refers to a nucleotide sequence, which can direct gene expression in cells. Such a promoter can be derived from a plant, a plant virus, or from any other living organism including bacteria and animals.

25 A plant promoter can be a constitutive promoter, such as, but not limited to, CaMV35S and CaMV19S promoters, FMV34S promoter, sugarcane bacilliform badnavirus promoter, CsVMV promoter, *Arabidopsis* ACT2/ACT8 actin promoter, *Arabidopsis* ubiquitin UBQ1 promoter, barley leaf thionin BTH6 promoter, and rice actin promoter.

The promoter can alternatively be a tissue specific promoter. Examples of plant tissue specific promoters include, without being limited to, bean phaseolin storage protein promoter, DLEC promoter, PHS β promoter, zein stprotein promoter, conglutin gamma promoter from soybean, AT2S1
5 gene promoter, ACT11 actin promoter from *Arabidopsis*, napA promoter from *Brassica napus*, potato patatin gene promoter and the Tob promoter.

The promoter may also be a promoter which is active in a specific developmental stage of a plant's life cycle, for example, a promoter active in late embryogenesis, such as: the LEA promoter; Endosperm-specific
10 expression promoter (the seed storage prolamin from rice is expressed in tobacco seed at the developmental stage about 20 days after flowering) or the promoter controlling the FbL2A gene during fiber wall synthesis stages.

In case of a tissue-specific promoter, it ensures that the heterologous protein is expressed only in the desired tissue, for example, only in the flower,
15 the fruit, the root, the seed, etc.

Both the tissue-specific and the non-specific promoters may be constitutive, i.e., may cause continuous expression of the heterologous protein.

The promoter may also be an inducible promoter, i.e., a promoter which is activated by the presence of an inducing agent, and only upon said
20 activation, causes expression of the heterologous protein. An inducing agent can be for example, light, chemicals, drought, high salinity, osmotic shock, oxidant conditions or in case of pathogenicity and include, without being limited to, the light-inducible promoter derived from the pea rbcS gene, the promoter from the alfalfa rbcS gene, the promoters DRE, MYC and MYB
25 active in drought; the promoters INT, INPS, prxEa, Ha hsp17.7G4 and RD21 active in high salinity and osmotic stress, the promoters hsr303J and str246C active in pathogenic stress, the copper-controllable gene expression system and the steroid-inducible gene system

Alternatively, an inducing agent may be an endogenous agent which is
30 normally present in only certain tissues of the plant, or is produced only at

certain time periods of the plant's life cycle, such as ethylene or steroids. By using such an endogenous tissue-specific inducing agent, it is possible to control the expression from such inducible promoters only in those specific tissues. By using an inducing agent produced only during a specific period of the life cycle, it is possible to control the expression from an inducible promoter to the specific phase in the life-cycle in which the inducing agent is produced.

Bacterial and yeast derived promoters are well known in the art.

Viruses are a unique class of infectious agents whose distinctive features are their simple organization and their mechanism of replication. In fact, a complete viral particle, or virion, may be regarded mainly as a block of genetic material (either DNA or RNA) capable of autonomous replication, surrounded by a protein coat and sometimes by an additional membranous envelope such as in the case of alpha viruses. The coat protects the virus from the environment and serves as a vehicle for transmission from one host cell to another.

Viruses that have been shown to be useful for the transformation of plant hosts include CaV, TMV and BV. Transformation of plants using plant viruses is described in U.S. Pat. No. 4,855,237 (BGV), EP-A 67,553 (TMV), Japanese Published Application No. 63-14693 (TMV), EPA 194,809 (BV), EPA 278,667 (BV); and Gluzman, Y. *et al.*, Communications in Molecular Biology: Viral Vectors, Cold Spring Harbor Laboratory, New York, pp. 172-189 (1988). Pseudovirus particles for use in expressing foreign DNA in many hosts, including plants, is described in WO 87/06261.

Construction of plant RNA viruses for the introduction and expression of non-viral foreign genes in plants is demonstrated by the above references as well as by Dawson, W. O. *et al.*, Virology (1989) 172:285-292; Takamatsu *et al.* EMBO J. (1987) 6:307-311; French *et al.* Science (1986) 231:1294-1297; and Takamatsu *et al.* FEBS Letters (1990) 269:73-76.

When the virus is a DNA virus, the constructions can be made to the virus itself. Alternatively, the virus can first be cloned into a bacterial plasmid for ease of constructing the desired viral vector with the foreign DNA. The virus n then be excised from the plasmid. If the virus is a DNA virus, a
5 bacterial origin of replication can be attached to the viral DNA, which is then replicated by the bacteria. Transcription and translation of this DNA will produce the coat protein which will encapsidate the viral DNA. If the virus is an RNA virus, the virus is generally cloned as a cDNA and inserted into a plasmid. The plasmid is then used to make all of the constructions. The RNA
10 virus is then produced by transcribing the viral sequence of the plasmid and translation of the viral genes to produce the coat protein(s) which encapsidate the viral RNA.

Construction of plant RNA viruses for the introduction and expression of non-viral foreign genes in plants is demonstrated by the above references as
15 well as in U.S. Pat. No. 5,316,931.

In one embodiment, a plant viral nucleic acid is provided in which the native coat protein coding sequence has been deleted from a viral nucleic acid, a non-native plant viral coat protein coding sequence and a non promoter, preferably the subgenomic promoter of the non-native coat protein coding
20 sequence, capable of expression in the plant host, packaging of the recombinant plant viral nucleic acid, and ensuring a systemic infection of the host by the recombinant plant viral nucleic acid, has been inserted. Alternatively, the coat protein gene may be inactivated by insertion of the non-native nucleic acid sequence within it, such that a protein is produced. The
25 recombinant plant viral nucleic acid may contain one or more additional non-native subgenomic promoters. Each non-native subgenomic promoter is capable of transcribing or expressing adjacent genes or nucleic acid sequences in the plant host and incapable of recombination with each other and with native subgenomic promoters. Non-native (foreign) nucleic acid sequences
30 may be inserted adjacent the native plant viral subgenomic promoter or the

native and a non-native plant viral subgenomic promoters if more than one nucleic acid sequence is included. The non-native nucleic acid sequences are transcribed or expressed in the host plant under control of the subgenomic promoter to produce the desired products.

5 In a second embodiment, a recombinant plant viral nucleic acid is provided as in the first embodiment except that the native coat protein coding sequence is placed adjacent one of the non-native coat protein subgenomic promoters instead of a non-native coat protein coding sequence.

10 In a third embodiment, a recombinant plant viral nucleic acid is provided in which the native coat protein gene is adjacent its subgenomic promoter and one or more non-native subgenomic promoters have been inserted into the viral nucleic acid. The inserted non-native subgenomic promoters are capable of transcribing or expressing adjacent genes in a plant host and are incapable of recombination with each other and with native
15 subgenomic promoters. Non-native nucleic acid sequences may be inserted adjacent the non-native subgenomic plant viral promoters such that said sequences are transcribed or expressed in the host plant under control of the subgenomic promoters to produce the desired product.

In a fourth embodiment, a recombinant plant viral nucleic acid is
20 provided as in the third embodiment except that the native coat protein coding sequence is replaced by a non-native coat protein coding sequence.

The viral vectors are encapsidated by the coat proteins encoded by the recombinant plant viral nucleic acid to produce a recombinant plant virus. The recombinant plant viral nucleic acid or recombinant plant virus is used to
25 infect appropriate host plants. The recombinant plant viral nucleic acid is capable of replication in the host, systemic spread in the host, and transcription or expression of foreign gene(s) in the host to produce the desired protein.

In many instances it is desired to target the expression of a recombinant protein. Such targeting can be into a cellular organelle or outside of the cell.
30 This can be effected, as is well known in the art, by appropriate signal

peptides, which are fused to the polypeptide to be targeted, typically at the N terminus.

Thus, as used herein and in the claims section which follows, the phrase "signal peptide" refers to a stretch of amino acids which is effective in targeting a protein expressed in a cell into a target location. Different signal peptides, which are known in the art, are effective in secreting a protein from bacteria, yeast, plant and animal cells.

It should be noted in this respect that signal peptides serve the function of translocation of produced protein across the endoplasmic reticulum membrane. Similarly, transmembrane segments halt translocation and provide anchoring of the protein to the plasma membrane, see, Johnson *et al.* The Plant Cell (1990) 2:525-532; Sauer *et al.* EMBO J. (1990) 9:3045-3050; Mueckler *et al.* Science (1985) 229:941-945. Mitochondrial, nuclear, chloroplast, or vacuolar signals target expressed protein correctly into the corresponding organelle through the secretory pathway, see, Von Heijne, Eur. J. Biochem. (1983) 133:17-21; Yon Heijne, J. Mol. Biol. (1986) 189:239-242; Iturriaga *et al.* The Plant Cell (1989) 1:381-390; McKnight *et al.*, Nucl. Acid Res. (1990) 18:4939-4943; Matsuoka and Nakamura, Proc. Natl. Acad. Sci. USA (1991) 88:834-838. A recent book by Cunningham and Porter (Recombinant proteins from plants, Eds. C. Cunningham and A.J.R. Porter, 1998 Humana Press Totowa, N.J.) describe methods for the production of recombinant proteins in plants and methods for targeting the proteins to different compartments in the plant cell. In particular, two chapters therein (14 and 15) describe different methods to introduce targeting sequences that results in accumulation of recombinant proteins in compartments such as ER, vacuole, plastid, nucleus and cytoplasm. The book by Cunningham and Porter is incorporated herein by reference. Presently, the preferred site of accumulation of the fusion protein according to the present invention is the ER using signal peptide such as Cel 1 or the rice amylase signal peptide at the N-terminus and an ER retaining peptide (HDEL, SEQ ID NO:17; or KDEL, SEQ ID NO:24) at the C-terminus.

According to an additional aspect of the present invention there is provided a method of producing recombinant β -glucosidase. The method according to this aspect of the present invention is effected by introducing, in an expressible or overexpressible form, a nucleic acid construct into a host cell. The nucleic acid construct includes a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger* and encoding a polypeptide having a β -glucosidase catalytic activity. The polynucleotide preferably further encodes a signal peptide in frame with the polypeptide. Still preferably, the polynucleotide further encodes an endoplasmic reticulum retaining peptide in frame with the polypeptide.

As used herein the term "introducing" refers both to transforming and to virally infecting, as these terms are further defined hereinabove. As used herein the terms "expressible form" and "overexpressible form" refers to a recombinant form which includes the required regulatory elements to effect expression or over expression of a coding region, all as is further detailed hereinabove.

According to a preferred embodiment of this aspect of the present invention, after sufficient expression has been detected, the polypeptide having the β -glucosidase catalytic activity is extracted from the expressing host cell.

Thus host cells, expressing the polypeptide according to the present invention, provide an immediate, easy and indefinite source of the polypeptide.

Any number of well-known liquid or solid culture media may be used for appropriately culturing host cells of the present invention, although growth on liquid media is preferred as the secretion of the polypeptide into the media results in simplification of polypeptide recovery. As is further detailed hereinabove, such secretion can be effected by the incorporation of a suitable signal peptide. The β -glucosidase may be isolated or separated or purified from host cell preparations using techniques well known in the art, such as, but not limited to, centrifugation filtration, chromatography, electrophoresis and

dialysis. Further concentration and/or purification of the β -glucosidase may be effected by use of conventional techniques, including, but not limited to, ultrafiltration, further dialysis, ion-exchange chromatography, HPLC, size-exclusion chromatography, cellobiose-sepharose affinity chromatography, and electrophoresis, such as polyacrylamide-gel-electrophoresis (PAGE). Using these techniques, β -glucosidase may be recovered in pure or substantially pure form.

According to an additional aspect of the present invention there is provided a method of increasing a level of at least one fermentation substance in a fermentation product. The method according to this aspect of the present invention is effected by fermenting a glucose containing fermentation starting material by a yeast cell overexpressing a nucleic acid construct which includes a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger* and which encodes a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the at least one fermentation substance in the fermentation product. The polynucleotide preferably further encodes a signal peptide in frame with the polypeptide. Still preferably, the polynucleotide further encodes an endoplasmic reticulum retaining peptide in frame with the polypeptide.

According an alternative aspect of the present invention there is provided a method of increasing a level of at least one fermentation substance in a fermentation product. The method according to this aspect of the present invention is effected by fermenting a plant derived glucose containing fermentation starting material by a yeast cell, the plant overexpressing a nucleic acid construct which includes a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger* and which encodes a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the at least one fermentation substance in the fermentation product. The polynucleotide preferably further encodes a signal peptide in frame with

the polypeptide. Still preferably, the polynucleotide further encodes an endoplasmic reticulum retaining peptide in frame with the polypeptide.

As used herein in the specification and in the claims section that follows, the term "fermentation" refers to a chemical change induced in a complex organic compound by the action of an enzyme, whereby the substance is split into simpler compounds. Specifically, the term "fermentation" includes the anaerobic dissimilation of substrates with the production of energy and reduced compounds, the final products thereof are organic acids, alcohols, such as ethanol, isopropanol, butanol, etc., and CO₂. Such products, are typically secreted and each of which is referred to herein as a "fermentation substance", i.e., any known fermentation resultant of either microbial or yeast fermentation.

As used herein in the specification and in the claims section that follows, the phrase "fermentation product" refers to the resultant material of a fermentation process. Examples include, but are not limited to, alcohol containing fermentation medium and alcoholic beverages, such a, but not limited to, fruit-based alcohol-containing beverages, wines and beers.

When used in conjunction with, for example, a β -glucanase, the β -glucosidase is effective for hydrolyzing a variety of cellulose containing materials to glucose. The glucose produced by enzymatic hydrolysis of the cellulose and other glucose containing saccharides, may be recovered and stored, or it may be subsequently fermented to ethanol using conventional techniques. Many processes for the fermentation of glucose generated from cellulose are well known, and are suitable for use herein. Briefly, the hydrolyzate containing the glucose from the enzymatic reaction is contacted with an appropriate microorganism under conditions effective for the fermentation of the glucose to ethanol. This fermentation may be separate from and follow the enzymatic hydrolysis of the cellulose (sequentially processed), or the hydrolysis and fermentation may be concurrent and conducted in the same vessel (simultaneously processed). Details of the various fermentation

techniques, conditions, and suitable microorganisms have been described, for example, by Wyman (1994, Bioresource Technol., 50:3-16) or Olsson and Hahn-Hagerdal (1996, Enzyme Microbial Technol., 18:312-331), the content of each of which is incorporated herein by reference. Following the
5 completion of a fermentation, the alcohol may be recovered by extraction, and optionally purified e.g., by distillation.

Thus, according to still another aspect of the present invention there is provided a method of producing an alcohol. The method according to this aspect of the present invention is effected by fermenting a glucose containing
10 fermentation starting material by a cell overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity, and extracting the alcohol therefrom. The polynucleotide preferably further encodes a signal peptide in frame with the polypeptide. Still
15 preferably, the polynucleotide further encodes an endoplasmic reticulum retaining peptide in frame with the polypeptide.

According to an additional aspect of the present invention there is provided a method of producing an alcohol. The method according to this aspect of the present invention is effected by fermenting a plant derived
20 glucose containing fermentation starting material by a cell, the plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity, and extracting the alcohol therefrom. The polynucleotide preferably further
25 encodes a signal peptide in frame with the polypeptide. Still preferably, the polynucleotide further encodes an endoplasmic reticulum retaining peptide in frame with the polypeptide.

Plants contain aroma and flavor compounds of glycosidic nature, their inherent aroma property can be released by degrading enzymes, turning a non-
30 volatile aroma compound into its volatile form. Thus, for example, α -L-

arabinofuranosidases, assist in the liberation of aroma compounds from substrates such as juices or wines, as described by Gunata et al. (European Patent Application No. 332.281, 1989; and "purification and some properties of an α -L-arabinofuranosidase from *A. niger* action on grape monoterpenyl arabinofuranosylglucosides. J. Agric. Food Chem. 38: 772-776, 1990). This outcome is achieved, for example, in a two step process wherein the first step comprises the use of an α -L-arabinofuranosidase, to catalyze the release of arabinose residues from monoterpenyl α -L-arabinofuranosyl glucosides contained in, for example, the fruit or vegetable juice via the cleavage of the (1
5 \rightarrow 6) linkage between a terminal arabinofuranosyl unit and the intermediate glucose of a monoterpenyl α -L-arabinofuranosylglucoside. The α -L-arabinofuranosidase is preferably in a purified form so as to avoid the undesirable degradation of other components of the juice which may be detrimental to its ultimate quality. In the second step, β -glucosidase is required
10 to yield the free terpenol from the resulting desarabinosylated monoterpenyl glucoside. If desired, both reaction steps may be performed in the same reaction vessel without the need to isolate the intermediate product (Gunata et al. (1989), supra). Thus, β -glucosidase is an essential contributor when the liberation of these aroma compounds for improving the flavor of the juice or
20 wine is desired. Moreover, in the case of wine, the control of the liberation of aroma compounds provides wines with a more consistent flavor, thus reducing or eliminating the undesirable effect of "poor vintage years". Additional information is contained in: "Cloning and expression of DNA molecules encoding arabinan degrading enzyme of fungal origin", U.S. Pat. No.
25 5,863,783; Y. Gueguen, et al. "A Very Efficient β -Glucosidase Catalyst for the Hydrolysis of Flavor Precursors of Wines and Fruit Juices", J. Agric. Food Chem. 44:2336-2340, 1996, each of which is incorporated herein by reference.

Thus, according to a further aspect of the present invention there is provided a method of increasing a level of at least one aroma substance in a

plant derived product, such as, but not limited to, an alcoholic beverage. The method according to this aspect of the present invention is effected by incubating a glucose containing plant starting material with a yeast cell overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger* which encodes a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the at least one aroma substance in the plant derived product. The polynucleotide preferably further encodes a signal peptide in frame with the polypeptide. Still preferably, the polynucleotide further encodes an endoplasmic reticulum retaining peptide in frame with the polypeptide.

While reducing the present invention to practice it was discovered that in order to obtain activity of a β -glucosidase in a transgenic plant, the expression construct should include a signal peptide. In addition, it was found that retaining the enzyme in the endoplasmic reticulum results in higher release of aroma compounds following homogenization and incubation. It is assumed that compartmentalization of the enzyme in for example the ER prevents it from interacting with its substrates which are mainly outside the cells, limiting such interaction following homogenization. Indeed, directing the enzyme to the apoplast resulted in increased release of aroma *in vivo*. Thus, depending on the specific application, one can chose weather to include in the construct an endoplasmic reticulum retaining peptide or not.

According to yet a further aspect of the present invention there is provided a method of increasing a level of at least one aroma substance in a plant derived product, such as, but not limited to, an alcoholic beverage. The method according to this aspect of the present invention is effected by incubating a glucose containing plant starting material with a yeast cell, said plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger* which encodes a polypeptide having a β -glucosidase

catalytic activity, thereby increasing the level of the at least one aroma substance in the plant derived product. The polynucleotide preferably further encodes a signal peptide in frame with the polypeptide. Still preferably, the polynucleotide further encodes an endoplasmic reticulum retaining peptide in
5 frame with the polypeptide.

As used herein in the specification and in the claims section that follows, the phrase "glucose containing starting material" refers to any source of energy, in the form of glucose containing compounds, other than free glucose, including, but not limited to, crushed, minced, diced or extracted plant
10 material, plant, or portions thereof, such as fruits, examples thereof are tropical fruits and grapes.

According to an additional aspect of the present invention there is provided a method of producing an aroma spreading plant. As used herein in the specification and in the claims section that follows, the phrase "aroma
15 spreading plant" refers to substantially any part of a plant, in which volatile compounds are generated by the catalytic activity of the β -glucosidase polypeptide of the present invention, release of volatile compounds therefrom is perceived by the olfactory system of an organism, such as a human.

The method according to this aspect of the present invention is effected
20 by overexpressing in the plant a nucleic acid construct including a genomic, complementary or composite polynucleotide derived from *Aspergillus niger*, which encodes a polypeptide having a β -glucosidase catalytic activity, thereby increasing aroma spread from the plant. Such overexpression is preferably performed in a tissue specific manner by, for example, employing a tissue
25 specific promoter, as hereinabove described, to thereby overexpress a heterologous protein in a selected portion of the plant. The tissue in which such overexpression is effected is selected according to the availability of glucose containing non-volatile aroma substrates therein. Thus, such an overexpression will cause the release of a volatile and aroma constituent of the
30 substrate. Thus, according to preferred embodiments overexpressing the

nucleic acid construct is limited to at least one tissue, such as a flower, a fruit, a seed, a root, a stem, pollen and leaves.

According to still a further aspect of the present invention there is provided a method of increasing a level of free glucose in a glucose containing fermentation starting material. The method according to this aspect of the present invention is effected by fermenting the glucose containing fermentation starting material by a cell overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, which encodes a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the free glucose in the glucose containing fermentation starting material. The polynucleotide preferably further encodes a signal peptide in frame with the polypeptide. Still preferably, the polynucleotide further encodes an endoplasmic reticulum retaining peptide in frame with the polypeptide.

According to another aspect of the present invention there is provided a method of increasing a level of free glucose in a plant derived glucose containing fermentation starting material. The method according to this aspect of the present invention is effected by fermenting the plant derived glucose containing fermentation starting material by a cell, the plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, which encodes a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the free glucose in the plant. The polynucleotide preferably further encodes a signal peptide in frame with the polypeptide. Still preferably, the polynucleotide further encodes an endoplasmic reticulum retaining peptide in frame with the polypeptide.

As used herein in the specification and in the claims section that follows, the term "free glucose" refers to glucose residues in the form of a monosaccharide, the levels of which are increased by the catalytic activity of β -glucosidase.

As used herein in the specification and in the claims section that follows, the phrase "glucose containing fermentation starting material" refers to any source of energy, in the form of glucose containing compounds, other than free glucose, including, but not limited to, crushed, minced, diced or
5 extracted plant material, plant, or portions thereof, used in industrial fermentation processes.

According to yet another aspect of the present invention there is provided a method of increasing a level of extra- or intracellular free glucose in a plant. The method according to this aspect of the present invention is
10 effected by overexpressing in the plant a nucleic acid construct including a genomic, complementary or composite polynucleotide preferably derived from *Aspergillus niger*, which encodes a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the free glucose in the plant. Thus, sweeter fruits can be produced. The polynucleotide preferably further
15 encodes a signal peptide in frame with the polypeptide. Still preferably, the polynucleotide further encodes an endoplasmic reticulum retaining peptide in frame with the polypeptide.

Glycosidases, including β -glucosidase, catalyze reactions involving the hydrolysis of O-glycosidic bond of glycosides, and synthesize oligosaccharides
20 when the reaction is run in reverse from the normal direction, a result achieved by, for example, site directed mutagenesis, and Km reversal. As described in the Background section hereinabove, the hydrolysis reaction mechanism of glycosidases involves two catalytic steps, the second of which involves a base catalyzed H_2O attack, resulting in the regeneration of the enzyme, and the
25 release of the saccharide residue. Thus, in addition, oligosaccharide synthesis can be achieved by adding a second saccharide to the reaction mixture, which competes with the H_2O molecule, and reacts in its place with the first saccharide in, what is known as, a transglycosylation reaction. Hence, as glycosidases are generally available and easy to handle, these enzymes have
30 the potential to catalyze the production of many different products using

inexpensive substrates. For further detail see U.S. pat. No. 5,716,812, which is incorporated herein by reference.

Thus, according to yet an additional aspect of the present invention there is provided a method of synthesizing oligosaccharides. The method according to this aspect of the present invention is effected by mixing a polypeptide having a β -glucosidase catalytic activity with first and second saccharide molecules to thereby join the first and second saccharide molecules into an oligosaccharide.

Additional objects, advantages, and novel features of the present invention will become apparent to one ordinarily skilled in the art upon examination of the following examples, which are not intended to be limiting. Additionally, each of the various embodiments and aspects of the present invention as delineated hereinabove and as claimed in the claims section below finds experimental support in the following examples.

15

EXAMPLES

Reference is now made to the following examples, which together with the above descriptions, illustrate the invention in a non limiting fashion.

Generally, the nomenclature used herein and the laboratory procedures utilized in the present invention include molecular, biochemical, microbiological and recombinant DNA techniques. Such techniques are thoroughly explained in the literature. See, for example, "Molecular Cloning: A laboratory Manual" Sambrook *et al.*, (1989); "Current Protocols in Molecular Biology" Volumes I-III Ausubel, R. M., ed. (1994); Ausubel *et al.*, "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Maryland (1989); Perbal, "A Practical Guide to Molecular Cloning", John Wiley & Sons, New York (1988); Watson *et al.*, "Recombinant DNA", Scientific American Books, New York; Birren *et al.* (eds) "Genome Analysis: A Laboratory Manual Series", Vols. 1-4, Cold Spring Harbor Laboratory Press, New York (1998); methodologies as set forth in U.S. Pat. Nos. 4,666,828; 4,683,202; 4,801,531; 5,192,659 and 5,272,057; "Cell Biology: A

Laboratory Handbook", Volumes I-III Cellis, J. E., ed. (1994); "Current Protocols in " Volumes I-III Coligan J. E., ed. (1994); Stites *et al.* (eds), "Basic and Clinical Immunology" (8th Edition), Appleton & Lange, Norwalk, CT (1994); Mishell and Shiigi (eds), "Selected Methods in Cellular Immunology", W. H. Freeman and Co., New York (1980); available immunoassays are extensively described in the patent and scientific liter, see, for example, U.S. Pat. Nos. 3,791,932; 3,839,153; 3,850,752; 3,850,578; 3,853,987; 3,867,517; 3,879,262; 3,901,654; 3,935,074; 3,984,533; 3,996,345; 4,034,074; 4,098,876; 4,879,219; 5,011,771 and 5,281,521; "Oligonucleotide Synthesis" Gait, M. J., ed. (1984); "Nucleic Acid Hybridization" Hames, B. D., and Higgins S. J., eds. (1985); "Transcription and Translation" Hames, B. D., and Higgins S. J., eds. (1984); "Animal Cell Culture" Freshney, R. I., ed. (1986); "Immobilized Cells and Enzymes" IRL Press, (1986); "A Practical Guide to Molecular Cloning" Perbal, B., (1984) and "Methods in Enzymology" Vol. 1-317, Academic Press; "PCR Protocols: A Guide To Methods And Application", Academic Press, San Diego, CA (1990); Marshak *et al.*, "Strategies for Protein Purification and Characterization - A Laboratory Course Manual" CSHL Press (1996); all of which are incorporated by reference as if fully set forth herein. Other general references are provided throughout this document. The procedures therein are believed to be well known in the art and are provided for the convenience of the reader. All the information contained therein is incorporated herein by reference.

MATERIALS AND EXPERIMENTAL METHODS

Purification of *A. niger* β -glucosidase:

A crude preparation of *A. niger* B1 (CMI CC 324626) β -glucosidase was obtained from Shaligal Ltd. (Tel-Aviv, Israel). A sample (10 ml) of the crude enzyme (140 Units/ml) was first diafiltered through a 50 kDa cut-off Amicon membrane (Amicon Corp., Danvers, MA), with 20 mM citrate buffer pH=5. The proteins were then separated on an FPLC equipped with a Mono-Q RH 5/5 column (Amersham Pharmacia Biotech AB, Uppsala, Sweden),

equilibrated with the same buffer. The enzyme was eluted with a linear gradient of 0 to 350 mM NaCl. Active fractions (see below, enzyme assays) were monitored and pooled (between 80-110 mM NaCl). The partially purified enzyme was dialyzed against 20 mM citrate buffer pH=3.5, applied to a Resource-S column equilibrated with the same buffer, and eluted with a gradient of 0-1 M NaCl. The purified enzyme (eluted at 155 mM NaCl) was concentrated by ultrafiltration (50 kDa cut-off membrane, Amicon).

Enzyme assays:

β -glucosidase enzyme activity was monitored using a plate assay as follows. 4-methylumbelliferone β -D-glucopyranoside (MUGlc, Sigma Chemical Inc. St. Louis, Missouri) to a final concentration of 0.5 mM, was dissolved in PC buffer (50 mM phosphate, 12 mM citric acid, pH=3.4) at 45 °C. The solution was mixed with 3 % agar in water, previously boiled and then cooled to 45 °C. The resulting solution (20 ml) was poured into a petri dish and after solidification, 10 μ l enzyme samples were spotted. The plate was incubated at 50 °C for one hour, and then illuminated with long UV. An intense fluorescence was indicative of β -glucosidase activity.

Detection of β -glucosidase in polyacrylamide gels was carried out by washing the SDS-polyacrylamide gel with 1:1 isopropanol:PC buffer to remove SDS and renature the enzyme. The gel was washed once in PC buffer and incubated in a thin layer of a solution of 0.5 mM MUGlc. After incubation at 50 °C for one hour, the active protein band was visualized by UV light.

Quantitative assays were performed using pNPGlc as a substrate according to Shoseyov (7).

Determination of thermal stability of *A. niger* β -glucosidase:

Recombinant enzyme (40 μ g/ml) was dissolved in 20 mM citrate phosphate buffer, pH=5. Each tested sample (8 μ l) was covered by 15 μ l mineral oil. The activity was determined by the standard pNPGlc assay (7).

Deglycosylation of *A. niger* β -glucosidase by N-glycosidase-F:

A N-glycosidase-F (Boehringer Mannheim, Mannheim, Germany) reaction mixture, containing 0.125 μ g pure β -glucosidase (previously denatured by boiling for 3 minutes in 1 % SDS and 5 % β -mercaptoethanol),
5 0.2 units of the N-glycosidase-F, sodium phosphate buffer (50 mM, pH=7.5), EDTA (25 mM), 1 % Triton X-100 and 0.02 % sodium azide, in a total volume of 12.5 μ l, was incubated for 4 hours at 37 °C. Reaction was stopped by addition of PAGE sample application buffer followed by 3 minutes of boiling.

Proteolysis and N-terminal sequences of *A. niger* B1 β -glucosidase:

10 Partial enzymatic proteolysis with *Staphylococcus aureus* V8 protease was carried out as described by Cleveland (28). Briefly, FPLC-purified β -glucosidase (5 μ g), was concentrated by acetone precipitation. The protein was separated on a preparative 10 % SDS-PAGE. The gel was stained with coomassie blue, destained and rinsed with cold water, and the β -glucosidase
15 protein band was excised. The resulting gel slice was applied to a second SDS-PAGE gel (15 % acrylamide) and overlaid with *Staphylococcus aureus* V8 protease. Digestion was carried out within the stacking gel by turning off the current for 30 min. As the bromophenol blue dye neared the bottom of the stacking gel, the current was restored. The electrophoresed cleavage products
20 were electroblotted to PVDF membranes. The native protein was transferred to PVDF in parallel. The N-terminal sequence of the native protein and two of the numerous cleavage products were analyzed by Edman degradation using a gas-phase protein sequencer (Applied Biosystems model 475A microsequencer).

Cloning of *bgl1* cDNA and genomic gene:

25 **Total RNA isolation:** Total RNA was isolated from *Aspergillus niger* B1 as follows: *A. niger* B1 was grown in liquid culture consisting of mineral media (NH₄)₂SO₄·3H₂O (0.5 g/l), KH₂PO₄ (0.2 g/l), MgSO₄ (0.2 g/l), CaCl₂·H₂O (0.1 g/l), FeSO₄·6H₂O (0.001 g/l), ZnSO₄·7H₂O (0.001 g/l), and

2 mM citric acid, at pH=3.5 with 1 % w/v bran as a carbon source. The medium was autoclaved, cooled and inoculated with *A. niger* B1 (10⁶ spores/ml). Baffled flasks were used with constant shaking (200 RPM) at 37 ° C. The appearance of β -glucosidase activity was monitored by placing 5 μ l of growth medium on 1 % agar plates containing 0.5 mM MUGlc, as described above. Activity was detected following 15 hours incubation. The mycelium was harvested following 24 hours growth period, and the medium removed by filtering through GFA glass microfibre (Whatman Inter. Ltd., Maidstone, England). The mycelium was then frozen with liquid nitrogen and ground to fine powder with mortar and pestle. Total RNA was produced from this powder by the Guanidine thiocyanate (TriReagent™) method (Molecular Research Center, Inc.).

RNA reverse-transcription reaction: cDNA was obtained by reverse transcribing total RNA (10 μ g) using Stratagene RT-PCR kit (Stratagene, La Jolla, CA). The reaction mixture (50 μ l) additionally consisted of: Oligo dT18 (1 μ g), RNase Block Ribonuclease Inhibitor (20 units), 1x buffer (50 mM Tris-HCl, pH=8.3, 75 mM KCl, 10 mM DTT, 3 mM MgCl₂), dNTPs (500 μ M each) and reverse transcriptase (300 units). Total RNA was initially denatured at 70 °C, allowed to cool to room temperature (for primers annealing), and added to the reaction mixture. The reaction mixture was incubated for 1 hour at 37 °C, followed by heating (95 °C, 5 minutes) and stored at -70 °C until further use.

DNA amplification: Degenerate primers for DNA amplification reaction by PCR methods were synthesized, based on part of the amino acid N-terminal sequence and an internal sequence, as determined by the Edman degradation, following V8 proteolysis (hereinbelow, experimental results). The partial sequence from β -glucosidase N-terminal derived amino acid sequence was Ser-Pro-Pro-Tyr-Tyr-Pro (SEQ ID NO:4), yielding the following primer: 5'-(C/G)(A/C/G/T)CC(A/C/G/T) CC(A/C/G/T)TA(C/T)TA(C/T)CC-3'

(SEQ ID NO:5). The partial sequence from E2 internal cleavage product amino acid sequence was Gln-Pro-Ile-Leu-Pro-Ala-Gly-Gly (SEQ ID NO:6), yielding the following primer: 5'-TCCIGC(T/G/C/A)GG(TG/C/A)A(G/A)(T/G/A)AT(T/G/C/A)GG(T/C)TG-3' (SEQ ID NO: 7).

5 DNA amplification reaction mixture (25 μ l) contained: reverse transcriptase reaction product (1 μ l), 10 x PCR buffer (2.5 μ l, Promega Corp., Madison, WI), dNTPs (250 μ M each), $MgCl_2$ (2.0 mM), degenerate primers (250 pmol each), DNA polymerase (3 units, Stratagene, La Jolla, CA) and overlaid with mineral oil (25 μ l). The reaction was performed in an automated
10 heating block (Programmable thermal controller - MJ Research, Inc.). PCR cycling conditions were 30 seconds denaturing at 94 $^{\circ}$ C, 60 seconds annealing at 50 $^{\circ}$ C, and 150 seconds elongation at 72 $^{\circ}$ C, repeated 36 times. The resulting amplified product was electrophoresed on a 1.2 % (w/v) agarose/TBE gel, resulting in a 2.2 kb cDNA gene fragment, which was further isolated
15 using Gel Extraction Kit (QIAGEN, Hilden, Germany) and cloned directly into the single 3'-T PCR insertion site of pGEM-T cloning vector (Promega Corp., Madison, WI).

Probe preparation: The 2.2 kb partial cDNA was digested with *Pst*I to produce a 1.2 kb fragment DNA probe. A sample (25 ng) of the fragment was
20 labeled with [32 P]dCTP, using the random sequence nanonucleotide rediprime DNA labeling system (Amersham Pharmacia Biotech AB, Buckinghamshire, England).

Preparation of genomic DNA plasmid library: An *A. niger* B1 genomic library was constructed in the pYEAUra3 yeast/*E. coli* shuttle vector
25 (Clontech Lab. Inc. Palo Alto, CA). *A. niger* B1 was grown in liquid culture as described above, the mycelium harvested following 48 hours of growth, frozen in liquid nitrogen and grounded. The mycelium ground was used to produce genomic DNA by the CTAB method of Murray and Thompson (29). The library was constructed from partially digested *Sau*3A genomic DNA,
30 cloned into the *Bam*HI site of the pYEUra3 yeast shuttle vector (Clontech Lab.

Inc. Palo Alto, CA). pYEAUra3 yeast/*E. coli* shuttle vector was digested with *Bam*HI and dephosphorylated with CIP to prevent self ligation. The partially digested genomic DNA was cloned into the shuttle vector with T4 ligase and used to transform TOP10 *E. coli* electro-competent cells, which were then
5 plated on LB-agar containing ampicillin (50 µg/ml). A total of 4×10^4 colonies were grown on LB-agar plates, blotted to Hybond-N membranes (Amersham Pharmacia Biotech AB, Buckinghamshire, England) and screened using the above described 1.2 kb probe. Positive clones were subcloned in pUC18 and sequenced (Biological Services, The Weizmann Institute of
10 Science, Rehovot, Israel).

Expression of bgl1 cDNA in E. coli:

Two specific primers were designed according to the 5' and the 3' sequences, corresponding to the N-terminal and C-terminal region of the mature protein: sense primer: 5'- (SEQ ID NO:8). Antisense primer: 5'-
15 AAAGGATCCTTAGTGAACAGTAGGCAGAGACGC-3' (SEQ ID NO:9). The isolated cDNA was digested with *Nco*I and *Bam*HI and cloned into a pET3d expression vector (Figure 1A, Novagen Inc., Madison, WI). Positive *E. coli* BL21(DE3) pLysS colonies, containing the *bgl1* cDNA, were confirmed by enzyme restriction and sequence analysis. Recombinant BGL1
20 was expressed according to the manufacturer's protocol.

Expression of bgl1 cDNA in Saccharomyces cerevisiae and Pichia pastoris:

The pYES2 vector (Invitrogen Inc., San Diego, CA) was used to successfully clone the *bgl1* cDNA gene into the *Hind*III/*Bam*HI of pYES2-
25 *bgl1* plasmid (Figure 1b), and transform *Saccharomyces cerevisiae* using the lithium acetate method (30). The BGL1 was expressed by inducing the Gal1 promoter according to the manufacturer's protocol. *Saccharomyces cerevisiae* strain INVSc2 (MATa, his3-D200, ura3-167) was used as the host. *Pichia pastoris* strain GS115 (his4 mutant) was used as the host for shuttle and
30 expression vector plasmid pHIL-S1 (Invitrogen Inc., San Diego, CA). The

bgl1 cDNA was cloned into the *EcoRI/BamHI* sites of pHIL-S1, yielding the pHIL-S1-bgl1 expression and secretion vector (Figure 1c). Expression in *P. pastoris* was carried out according to the manufacturer's protocol. Screening of β -glucosidase-expressing clones was facilitated by top-agar, containing 50 mg X-Glc, 30 ml methanol and 1 % agar per liter. Blue color indicated a colony producing active β -glucosidase.

Western blot analysis:

Antibodies were produced from rabbit serum 36 days following a second injection of 100 μ g purified protein and adjuvant (AniLab Biological Services, Tal-Sachar, Israel). High molecular weight ladder was from Sigma Chemical Inc. St. Louis, Missouri. Western blot conditions were as described in reference 36.

Determination of the Stereochemical Course of Hydrolysis:

The method was essentially as described by Wong *et al.* (31). PNPGlc (10 μ mol) was dissolved in 0.5 ml of 25 mM acetate buffer pH=3.5 in D₂O in an NMR tube. β -Glucosidase was lyophilized and redissolved in 100 μ l D₂O (35 units/ml). The ¹H-NMR spectrum of the substrate was recorded, enzyme added (10 μ l), and spectra recorded at specified time intervals on a Bruker AMX400 at 25 °C.

Inactivation and reactivation studies:

Pure *A. niger* β -Glucosidase enzyme (0.47 mg/ml) was incubated in the presence of various concentrations of 2-deoxy-2-fluoro- β -glucosyl fluoride (2FGlcF, 0.5-6 mM) in 30 mM citrate buffer pH=4.8 at 50 °C. Residual enzyme activity was determined at different time intervals by addition of an aliquot (10 μ l) of the inactivation mixture, to a solution containing citrate buffer (30 mM, pH=4.8), BSA (8 μ g) and 2,4-dinitrophenyl β -D-glucopyranoside (DNPGlc, 0.625 mM, 830 μ l). Release of DNP was determined spectrophotometrically by measuring the absorbance at 400 nm one minute after the addition of the substrate.

Reactivation rates were determined as follows: pure *A. niger* β -glucosidase (0.34 mg/ml) was preincubated with 2FGlcF (5 mM) for 15 min, after which the excess of the inactivator was diafiltered by 20-kDa nominal molecular mass cutoff centrifugal concentrators (Sartorius Inc., Goettingen, Germany). Samples of the purified, inactivated enzyme were incubated in the presence linamarin (0-16 mM) in citrate buffer (30 mM, pH=4.8) at 50 °C for 0, 10, 20 and 30 minutes, and the activity of each sample was determined using p-nitrophenyl β -D-glucopyranoside (pNPGlc) as a substrate.

Expression of bgl1 cDNA in tobacco plants:

Genetic constructs:

Bgl1 cDNA was cloned in pETB1 (37). pJD330 and pBINPlus (38) were used as an intermediate and binary vector, respectively. Cell1 signal sequence as well as 35S plus Ω fragment were retrieved from pB21, modified pBluescript SK (39). *Nicotiana tabacum* cv. Samson was used as a model plant for gene transformation. Three gene constructs were employed (Figures 11a-c): (i) bgl1 without any signal peptide which served for cytoplasmic expression (Figure 11a, plasmid pJDB1); (ii) bgl1 including a cell1 signal peptide at the N terminus for secretion into the apoplast (Figure 11b, plasmid pJDCB1); and (iii) bgl1 including the cell1 signal peptide and the KDEL (SEQ ID NO:24) ER-retaining peptide at the C-terminus for accumulation in the ER (Figure 11c, plasmid pJDCB1T).

To this end, bgl1 cDNA (2.5 kb) was released from pETB1 (37) with *Nco*I and *Bam*HI and inserted into pJD330 between the 35S promoter Ω fragment and the *nos* terminator, eliminating the *gus* gene, resulting in plasmid pJDB1. Endoplasmic reticulum retaining signal tetrapeptide HDEL (SEQ ID NO:17) was synthesized and fused with *bgl1* at the C-terminal in pJDB1 by a fidelity PCR reaction with the following pair of primers: Forward primer (23 mer), starting from nucleotide 1248 of bgl1 cDNA 5' - (1248) - CAGTGACCGTGGATGCGACAATG - (1270') - 3' (SEQ ID NO:20); Reverse primer (40 mer), starting at nucleotide 2506 of bgl1 cDNA encoding

also for the HDEL (SEQ ID NO:17) peptide 5' - (2506) -
AGAGACGGATGACAAGTACTACTTGAAATTGGGCCCAAAA-3' (SEQ
ID NO:21). For pJDCB1T (35S Ω + Cell + bgl1 + HDEL), the 35S Ω
fragment of pJDB1 was replaced by a 35S Ω + Cell fragment digested from
5 pB21 with *Bam*HI and *Xba*I. For pJDCB1 (35S Ω + Cell + bgl1), the
fragment containing 35S Ω and Cell as well as part of bgl1 was cut from
pJDCB1T with *Hind*III and *Nru*I and ligated with the vector of pJDB1
digested with the same pair of restriction enzymes. The nucleotide sequence
of all of the genetic constructs was confirmed by DNA sequencing.

10 Gene cassettes in the intermediate vectors of pJDB1, pJDCB1 and
pJDCB1T were further isolated with *Hind*III and *Eco*RI and inserted into
multiple cloning sites of the binary vector pBINPlus. Disarmed *Agrobacterium*
LB4404 was transformed with pBINPlus containing bgl1 gene cassettes.

Tobacco plant transformation:

15 The young leaves of *in vitro* grown plantlets were excised and cut into
0.5 cm² pieces and then immersed for 5 minutes in an overnight grown culture
of *Agrobacterium*. After blotted with sterile Whatman filter paper, the
infected leaves were co-cultured for 2 days with *Agrobacterium* on MS
medium plus 2.0 mg/L of Zeatin and 0.1 mg/L of IAA as well as 0.35 % (w/v)
20 phytagel and then transferred to the same medium but with 300 mg/L
kanamycin and 300 mg/L carbenicillin. Regenerates were then transferred to
the rooting media, containing only MS salts, vitamins and the same antibiotics.
Rooted plants were transferred to greenhouse after PCR screening.

Screening for transgenic plants:

25 DNA and protein of plants were extracted according to Nagy *et al.* (40).
PCR verification of gene insertion into plant genome was done with the
following pairs of primers, which cover the DNA fragment from position 1248
to the end of bgl1: 5'-CAGTGACCGTGGATGCGACAATG-3' (SEQ ID
NO:22) and 5'-AAAGGATCCTTAGTGAAACAGTAGGCAGAGACGC-3'
30 (SEQ ID NO:23).

Identifying transgenic plants expressing BGL1 protein and activity:

Western blot (40) and SDS-PAGE activity gel staining (37) were employed to screen successful transgenic lines, using the purified *A. niger* BGL1 protein as positive controls and non-transgenic plant as negative control.

SPMI-GC/MS analysis:

The effect of bgl1 on flavor compound evolution and composition was studied. Fresh leaves of transgenic plants and of wild type control plants were excised and ground in liquid nitrogen. Ice-cold extraction buffer, containing 10 mM EDTA, 4 mM DTT in 50 mM phosphate buffer, pH 4.3, was added in a ratio of 1:3 w/w. The mixture was then shaken for 0.5 hours. 0.75 ml of supernatant from each of the centrifuged mixtures was taken into a glass vial. All manipulations were at 4 °C. After 9 hours of incubation at 37 °C, the volatiles in the vial were analyzed according to Clark *et al.* (41) using a Saturn Varian 3800 SPMI-GC-MS apparatus, equipped with a DB-5 capillary column. The temperature of splitless injections was 250 °C and the transfer line was maintained at 280 °C. Helium was used as a carrier gas. The oven was programmed as follows: 1 minute at 40 °C with gradually heating to 250 °C at a rate of 5 °C/minute.

EXPERIMENTAL RESULTS***Purification of wild type A. niger β -glucosidase:***

A. niger β -glucosidase enzyme preparation was purified by Mono-Q FPLC. Active protein samples eluted from the Mono-Q column were separated on a 10 % SDS-PAGE gel, stained with coomassie blue, and incubated in the presence of MUGlc to demonstrate activity of the enzyme. At this stage of purification, a discrete band, having an apparent molecular mass of approximately 160 kDa and β -glucosidase activity could be detected (Figure 2b, lanes 1-5: 1 - electroeluted band of BGL1 from preparative PAGE-SDS gel stabs; 2-5 - acetone precipitates from Mono-Q separation of BGL1). However,

the apparent mass of the denatured enzyme (boiled for 10 min in the presence of β -mercaptoethanol), was shown to be 120 kDa on 10 % SDS-PAGE (Figure 2a). The enzyme was designated BGL1 was further purified to homogeneity on a Resource-S column (Figure 3). Deglycosylation of *A. niger* β -glucosidase was performed by N-glycosidase-F. As demonstrated in Figure 4, SDS-PAGE analysis indicated that approximately 20 kDa of the *A. niger* β -glucosidase mass can be attributed to N-linked carbohydrates.

Proteolysis and N-terminal sequences of BGL1:

Partial enzymatic proteolysis with *Staphylococcus aureus* V8 protease of purified BGL1 was conducted. The undigested protein and cleavage products were separated by SDS-PAGE, followed by electroblotting onto PVDF membranes and determination of the N-terminal sequence of the native protein and two of the cleavage products. Amino acid sequences obtained were as follows:

N-terminal native protein: Asp-Glu-Leu-Ala-Tyr-Ser-Pro-Pro-Tyr-Tyr-Pro-Ser-Pro-Trp-Ala-Asn-Gly-Gln-Gly-Asp (SEQ ID NO:10). Underlined portion represents SEQ ID NO:4.

Internal cleavage product - E1 polypeptide: Val-Leu-Lys-His-Lys-Asn-Gly-Val-Phe-Thr-Ala-Thr-Asp-Asn-Trp-Ala-Ile-Asp-Gln-Ile-Glu-Ala-Leu-Ala-Lys (SEQ ID NO: 11).

Internal cleavage product - E2 polypeptide: Gly-Ala-Thr-Asp-Gly-Ser-Ala-Gln-Pro-Ile-Leu-Pro-Ala-Gly-Gly-Gly-Pro-Gly-Gly-Asn-Pro (SEQ ID NO:12). Underlined portion represents SEQ ID NO:6.

FastA analysis (32) indicated that the N-terminal sequence, as well as the internal sequences, have sequence similarity with sequences of β -glucosidase from the yeast *Saccharomycopsis fibuligera* which belonging to Family 3 of the glycosyl hydrolases.

Isolation and characterization of bgl1 cDNA and genomic DNA:

In order to clone the *A. niger* β -glucosidase gene, degenerate primers were designed according to the sequence of digest fragments of the

polypeptide. These oligonucleotides were used to amplify a cDNA fragment of the β -glucosidase gene by RT-PCR. A 1.2 kb probe was excised from the resultant 2.2 kb amplification product and was used to screen a genomic library, constructed in pYEUra3 yeast/*E. coli* shuttle vector. Positive clones were successfully subcloned and sequenced, resulting in full length bgl1 genomic sequence (SEQ ID NO:3, Figure 5a). Amplification primers were then generated, according to the genomic DNA sequence, corresponding to the N- and C-terminal of the mature protein. RT-PCR was thereafter used for amplifying the full length β -glucosidase cDNA sequence (SEQ ID NO:1, Figure 5a, GenBank Accession No. AJ132386). The cDNA sequence perfectly matched the DNA sequence of the combined exons. The open reading frame was found to encode a polypeptide with a predicted molecular weight of 92 kDa. The gene includes 7 exons intercepted by 6 introns (Figure 5b). Analysis of the DNA sequence upstream to the sequence encoding for the mature protein revealed a putative leader sequence, intercepted by an 82 bp intron.

Production of rBGL1 in E. coli:

Recombinant BGL1 was overexpressed in *E. coli*. No apparent β -glucosidase activity could be detected in the *E. coli* extracts, however SDS-PAGE analysis revealed a relatively intense protein band expressed at the expected molecular weight. Western blot analysis using rabbit polyclonal anti-native BGL1 antibodies (AniLab Biological Services, Tal-Sachar, Israel), positively identified the 90 kDa protein band (not shown). Further analysis revealed that the protein was accumulated in inclusion bodies. Several refolding experiments were conducted, however, these efforts to produce active protein from *E. coli* failed (not shown).

Expression of recombinant BGL1 in S. cerevisiae and P. pastoris:

Recombinant BGL1 was successfully expressed both in *S. cerevisiae* and *P. pastoris*. In *S. cerevisiae* a relatively low level of expression was found. The recombinant protein was detected by a Western blot analysis

(Figure 6a). The total protein extract of *S. cerevisiae* expressing bgl1 cDNA had a β -glucosidase activity of 1.9 units/mg protein. No β -glucosidase activity was detected in control *S. cerevisiae*, transformed with vector only, under the same assay conditions. However, no protein band corresponding to recombinant BGL1 could be detected by coomassie blue staining. *P. pastoris* transformed with bgl1 secreted relatively high levels of recombinant BGL1 to the medium (about 0.5 g/l) appearing as an almost pure protein in the culture supernatant (Figure 6b). This recombinant enzyme was highly active (124 units/mg protein) and without further purification, yielded specific activity similar to that of the pure native enzyme.

¹H-NMR determination of stereochemical outcome:

¹H-NMR spectra of a reaction mixture containing pNPGlc and BGL1 revealed that the beta anomer of glucose was formed first (H-1 = 4.95 ppm), with delayed appearance of the alpha anomer (H-1 5.59 ppm), the consequence of mutarotation (Figure 7). BGL1 is indeed, therefore, a retaining glycosidase, as has been observed for other family members (33, 34).

Inactivation and reactivation of *A. niger* β -glucosidase:

Enzyme was incubated in the presence of various concentrations of 2FGlcF and residual enzyme activity was monitored at different time intervals. Enzyme activity decreased in a time-dependent manner, according to pseudo-first order kinetics, allowing the determination of pseudo-first order rate constants: $K_i = 4.5 \text{ min}^{-1}$ and $K_i = 35.4 \text{ mM}$, for inactivation at each inactivator concentration (0, 0.5, 1, 2, 4, and 6 mM, Figure 8).

Rates of reactivation of 2-deoxy-2-fluoroglucosyl-BGL1 were determined in the presence of different concentrations of linamarin by monitoring activity regain after 0, 10, 20 and 30 min (Figure 9). The regain of activity followed a first order process at each linamarin concentration.

Thermal stability of *A. niger* β -glucosidase:

Thermal stability of the recombinant enzyme was evaluated at different temperatures, presented as percent enzymatic activity relative to an enzyme

solution kept at 4 °C. Results obtained are summarized in Table 2 and illustrated in Figure 10. The purified enzyme exhibits high thermal stability, as majority (above 50 %) of the activity is maintained at a temperature ranging from 4-60 °C.

5

TABLE 2

Temp. °C	% activity
4	100
50	91.5
55	83.5
60	68
65	17.8

Expression of BGL1 in tobacco plants:

Agrobacterium mediated leaf disc transformation resulted in transgenic tobacco plants as was proved by PCR (Figure 12) for the presence of the transgene, Western blotting (Figures 13a-b) for presence of the protein and activity assays (Figures 14 and 15) for presence of protein activity. Table 3 below summarizes the results.

TABLE 3

Gene construct	BGL1	Cell + BGL1 + HDEL	Cell + BGL1
Number of Regenerates	33	14	27
PCR positive	29	9	23
Western Blot positive	4	9	18
Activity gel positive	0	9	18

15

Of the 29 PCR positive regenerates transformed with cDNA encoding BGL1, which fails to encode a signal peptide, only in 4 the BGL1 protein was detectable via Western blotting, however no BGL1 activity was measurable in any of which. The BGL1 was found smaller in molecular weight compared to wild type *A. niger* beta-glucosidase and of processed recombinant BGL1

containing a signal peptide. Its apparent size of about 95 kDa is very close to 92 kDa which is the calculated molecular weight of the un-glycosylated *A. niger* beta-glucosidase. This result coincides with the fact that a protein with no signal peptide is expected to be released from the ribosomes and remain in the cytoplasm (42) un-glycosylated, as protein glycosylation is conducted in the lumen of the endoplasmic reticulum (43).

Of the 9 PCR positive regenerates transformed with a cDNA encoding the BGL1 and a Cell signal peptide and in addition encodes the HDEL ER retaining peptide, all plants expressed detectable amounts of BGL1 protein and activity.

Of the 23 PCR positive regenerates transformed with a cDNA which encodes the BGL1 protein and the Cell signal peptide but not the HDEL ER retaining peptide, 18 plants expressed detectable amounts of BGL1 protein and activity.

The effect of BGL1 on flavor compound evolution and composition in transgenic tobacco plants:

Extracts of transgenic plants (CB14 and CBT21 containing similar BGL1 activity, see Figure 15) were incubated for 9 hours at 37 °C, and flavor compounds were analyzed by SPME-GC/MS. The results, which are summarized in Table 4 below, show that with the exception of oleyl alcohol, the concentration of different flavor compounds is increased in transgenic plants expressing active BGL1 compared with the control. Furthermore, it seems that compartmentalization of BGL1 in the ER (or for that matter, any other subcellular organelle), rather than its secretion to the apoplast, results in higher release of flavor compounds. It is likely that this is resulted from the localization many flavor compounds in the apoplast, thus, secretion of BGL1 to the apoplast cause *in vivo* release of flavor compounds, while compartmentalization of BGL1 in the ER results in release of flavor compounds only in the event of cell disruption and decompartmentalization.

TABLE 4

Retention Time (minutes)	Scan	Name	CB14	CBT 21
3.917	419	Hexanal	- ^a	-
4.749	508	3-methyl-pentanoic acid	-	-
4.863	520	2-Hexenal	-	+ ^b
5.167	552	?	-	+
6.564	702	1-Heptanol	-	-
7.1	752	?	+	++ ^d
8.085	865	2-ethyl-1-pexanol	-	+
8.132	870	Limonene	++	+
8.194	877	2-methyl-phenol	-	+
10.653	1139	Menthol	+	+
11.757	1258	Nerol	-	+
12.039	1288	6-Quinolinal	-	+
12.1	1294	2-butyl-1-octanol	-	+
13.0	1458	?	-	+
13.7	1466	?	-	+
14.091	1507	Vitispirane	-	+
14.094	1516	4-[2,6,6-trimethyl-1-cyclohexen-1-yl] 3-Buten-1-one	+	++
15.985	1710	?	-	-
19.327	2069	Oleyl alcohol	-- ^c	--

CB14 - transgenic plant containing Cell signal peptide + BGL1; CBT 21 - transgenic plant containing Cell signal peptide + BGL1 + HDEL ER retaining peptide. a - "-" means no significant difference in concentration compared with wile type. b - "+" means significant increase compared with the wild type. c - "--" means significant decrease compared with the wild type. d - "++" means significant increase compared with a respective mark "+". ? - unknown compound.

Although the invention has been described in conjunction with specific embodiments thereof, it is evident that many alternatives, modifications and variations will be apparent to those skilled in the art. Accordingly, it is intended to embrace all such alternatives, modifications and variations that fall within the spirit and broad scope of the appended claims. All publications, patents, patent applications and sequences identified by GenBank accession numbers mentioned in this specification are herein incorporated in their entirety by reference into the specification, to the same extent as if each

individual publication, patent, patent application or sequence was specifically and individually indicated to be incorporated herein by reference. In addition, citation or identification of any reference in this application shall not be construed as an admission that such reference is available as prior art to the
5 present invention.

REFERENCES:

1. Bause, E., and Legler, G. (1974) Hoppe-Seyler's Z. Physiol. Chem. 355, 438-442.
2. Beguin, P., and Aubert, J. P. (1994) FEMS Microbiol. Rev. 13, 25-58.
3. Rombouts, F. M. and Pilnik, W. (1978), Proc. Biochem. 8, 9-13.
4. Sternberg, D., Vijayakumar, P. and Reese, E. T. (1977), Can. J. Microbiol. 23, 139-147.
5. Woodward, J. and Wiseman, A. (1982), Enz. Microbiol. Technol. 4, 73-79.
6. Kerns, G., Dalchow, E., Klappach, G. and Meyer, D. (1986), Acta. Biotechnol. 6(4), 355-359.
7. Shoseyov, O., Bravdo, B., Ikan, R. and Chet, I. (1988), Phytochem. 27(7), 1973-1976.
8. Shoseyov, O., Bravdo, B., Siegel, D., Goldman, A., Cohen, S. and Shoseyov, L. (1990), J. Agric. Food. Chem. 39, 1387-1390.
9. Dekker, R. F. H. (1986), Biotechnol. Bioengin. 26, 1438-1442.
10. Kitpreechavanich, V. M., Hayashi, M. and Nagai, S. (1986), Agric. Biol. Chem. 50, 1703-1711.
11. Yeoh, H. H., Tan, T. K. and Koh, S. K. (1986), Appl. Microbiol. Biotechnol. 25, 25-28.
12. Crouzet, J. and Chassagne, D. (1999) in Naturally Occurring Glycosides (Ikan. R., ed.) pp.225-274, Wiley Press.
13. Prade, H., L.F. Mackenzie and S.G. Withers (1998) *Carbohydr. Res.* 305, 371-381.
14. Zarevucka M., Vacek, M., Wimmer, Z., Demnerova, K. and Mackova, M. (1998) *Chirality* 10, 676-68.
15. Yi, Q., Sarney, D.B., Khan, J.A. and Vulfson, E.N. (1998) *Biotechnol. Bioeng.* 60, 385-390.

16. McCleary, B. V. and Harrington, J. (1988), *Methods Enzymology*, 160, 575-583.
17. Watanabe, T., Sato, T., Yoshioka, S., Koshijima, T. and Kuwahara, M. (1992), *Eur. J. Biochem.* 209, 651-659.
18. Unno, T., Ide, K., Yazaki, T., Tanaka, Y., Nakakuki, T. and Okada, G. (1993), *Biosci. Biotech. Biochem.* 57(12), 2172-2173.
19. Le Traon, M. M. P., and Pellerin P., (1998) *Enz. Micro. Technol.*, 22 (5) 374-382.
20. Penttila, M. E., Nevalainen, H. K. M., Raynal, A., and Knowles, J. K. C., (1984), *Mol. Gen. Genet.* 194, 494-499.
21. Henrissat, B., and Bairoch, A. (1996) *Biochem. J. Lett.* 316, 695-696
22. Sinnott, M. L. (1990) *Chem. Rev.* 90, 1171-1202.
23. McCarter, J., and Withers, S. G. (1994) *Curr. Opin. Struct. Biol.* 4, 885-892.
24. Davies, G., Sinnott, M. L., and Withers, S. G. (1998) in *CBiological Catalysis* (Sinnott, M. L., ed.) Vol. 1, pp. 119-208, Academic Press.
25. Legler, G., Roeser, K. R., and Illig, H. K. (1979) *Eur. J. Biochem.* 101, 85-92.
26. Roeser, K. R., and Legler, G. (1981) *Biochem. Biophys. Acta.* 657, 321-333.
27. Legler, G., Sinnott, M. L., and Withers, S. G. (1980) *J. Chem. Soc., Perkin Trans. II*, 1376-1383.
28. Cleveland, D. W., Fischer, S. G., Kirschner, M. W. and Laemmli, U.K. (1977), *J. Biol. Chem.* 252(3), 1102-1106.
29. Murray, M. G. and Thompson, W. F. (1980), *Nucl. Acids Res.* 8, 4321-4325.
30. Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983), *J. Bacteriol.* 153(1), 163-168.
31. Wong, A.W., He, S., Grubb, J. H., Sly, W. S. and Withers S. G. (1998), *J. Biol. Chem.* 273, 34057-34062.

32. Lipman, D. J. and Pearson, W. R. (1985) *Science*, 227, 1435-1441.
33. Legler, G. (1968) Hoppe-Zeyler's *Z. Physiol. Chem.* 349, 767-774.
34. Hrmova, M. and Fincher, G. B. (1996) *J. Biol. Chem.* 271, 5277-5286.
35. Iwashita K., Todoroki, M., Kimura, H., Shimoi, H. and Ito, K. (1998) *Biosci. Biotechnol. and Biochem.* 62 (10), 1938-1946.
36. Rlow E. and Lane D. (1988) *Antibodies, a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
37. Siegel D., Marton I., Dekel M., Bravdo B.A., He S.M., Wither S.G., Shoseyov O., 2000, Cloning, expression, characterization, and nucleophile identification of family 3, *Aspergillus niger* beta-glucosidase. *J. Biol. Chem.* 275(7),4973-4980.
38. Engelen F.A.V., Molthoff J. W., Conner A. J., Nap J. P., Pereira A and Stiekema W., 1995, pBINPLUS: an improved plant transformation vector based on pBIN19. *Transgenic Res.* 4, 288-290.
39. Shani Z., Dekel M., Tsabary G., Shoseyov., 1997, Cloning and characterization of elongation specific endo-1,4- beta-glucanase (Cell) from *Arabidopsis thaliana*, *Plant Molecular Biol.* 34: 837-842.
40. Nagy F., Kay S.A.K., Chua N. H., 1989, Analysis of gene expression in transgenic plants. In *Plant Molecular Biology: manual* edited by Gelvin et al., Kluwer Academic Publishers.
41. Clark T. J., Bunch J. E., 1997, Qualitative and quantitative analysis of flavor additives on tobacco products using SPME-GC-Mass Spectroscopy. *J. Agric. Food Chem.* 45 (3),844-849.
42. Lewin B., 1994, The apparatus for protein localization, In *Gene V*, Oxford University Press. Pp:279-314.
43. Prodi J. A., 2000, The role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *Biochem. J.*, 348, 1-13.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid comprising a genomic, complementary or composite polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity.
2. The isolated nucleic acid of claim 1, wherein said polynucleotide is as set forth in SEQ ID NOs:1, 3 or a portion thereof.
3. A nucleic acid construct comprising the isolated nucleic acid of claim 1.
4. The nucleic acid construct of claim 3, further comprising at least one cis acting control element for regulating expression of said polynucleotide.
5. A host cell comprising the nucleic acid construct of claim 3.
6. The host cell of claim 5, wherein the host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.
7. The host cell of claim 6, wherein said prokaryotic cell is *E. coli*.
8. The host cell of claim 6, wherein said eukaryotic cell is selected from the group consisting of a yeast cell, a fungous cell, a plant cell and an animal cell.
9. An organism comprising the nucleic acid construct of claim 3.
10. The organism of claim 9, wherein the organism is a plant.

11. A recombinant protein comprising an *Aspergillus niger* derived polypeptide having a β -glucosidase catalytic activity.
12. The recombinant protein of claim 11, wherein said polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.
13. A method of producing recombinant β -glucosidase, the method comprising the step of introducing, in an expressible form, a nucleic acid construct into a host cell, said nucleic acid construct including a genomic, complementary or composite polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity.
14. The method of claim 13, further comprising the step of extracting said polypeptide having said β -glucosidase catalytic activity.
15. The method of claim 13, wherein said host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.
16. The method of claim 15, wherein said prokaryotic cell is *E. coli*.
17. The method of claim 15, wherein said eukaryotic cell is selected from the group consisting of a yeast cell, a fungous cell, a plant cell and an animal cell.
18. The method of claim 13, wherein said polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.

19. The method of claim 13, wherein said polynucleotide is as set forth in SEQ ID NOs: 1, 3 or a portion thereof.
20. A method of producing a recombinant β -glucosidase overexpressing cell, the method comprising the step of introducing, in an overexpressible form, a nucleic acid construct into a host cell, said nucleic acid construct including a genomic, complementary or composite polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity.
21. The method of claim 20, wherein said host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.
22. The method of claim 21, wherein said prokaryotic cell is *E. coli*.
23. The method of claim 21, wherein said eukaryotic cell is selected from the group consisting of a yeast cell, a fungus cell, a plant cell and an animal cell.
24. The method of claim 20, wherein said polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.
25. The method of claim 20, wherein said polynucleotide is as set forth in SEQ ID NOs: 1, 3 or a portion thereof.
26. A method of increasing a level of at least one fermentation substance in a fermentation product, the method comprising the step of fermenting a glucose containing fermentation starting material by a yeast cell overexpressing a nucleic acid construct including a genomic, complementary

or composite polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the at least one fermentation substance in the fermentation product.

27. The method of claim 26, wherein said polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.

28. The method of claim 26, wherein said polynucleotide is as set forth in SEQ ID NOs: 1, 3 or a portion thereof.

29. A method of increasing a level of at least one fermentation substance in a fermentation product, the method comprising the step of fermenting a plant derived glucose containing fermentation starting material by a yeast cell, said plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the at least one fermentation substance in the fermentation product.

30. The method of claim 29, wherein said polypeptide as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.

31. The method of claim 29, wherein said polynucleotide is as set forth in SEQ ID NOs: 1, 3 or a portion thereof.

32. A method of increasing a level of at least one aroma substance in a plant derived product, the method comprising the step of incubating a glucose containing plant starting material with a yeast cell overexpressing a nucleic acid construct including a genomic, complementary or composite

polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the at least one aroma substance in the plant derived product.

33. The method of claim 32, wherein said plant derived product is a fermentation product.

34. The method of claim 32, wherein said polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.

35. The method of claim 32, wherein said polynucleotide is as set forth in SEQ ID NOs: 1, 3 or a portion thereof.

36. A method of increasing a level of at least one aroma substance in a plant derived product, the method comprising the step of incubating a glucose containing plant starting material with a yeast cell, said plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the at least one aroma substance in the plant derived product.

37. The method of claim 36, wherein said plant derived product is a fermentation product.

38. The method of claim 36, wherein said polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.

39. The method of claim 36, wherein said polynucleotide is as set forth in SEQ ID NOs: 1, 3 or a portion thereof.

40. A method of increasing a level of free glucose in a glucose containing fermentation starting material, the method comprising the step of fermenting the glucose containing fermentation starting material by a cell overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the free glucose in the glucose containing fermentation starting material.

41. The method of claim 40, wherein said cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.

42. The method of claim 41, wherein said prokaryotic cell is *E. coli*.

43. The method of claim 41, wherein said eukaryotic cell is selected from the group consisting of a yeast cell, a fungous cell, a plant cell and an animal cell.

44. The method of claim 40, wherein said polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.

45. The method of claim 40, wherein said polynucleotide is as set forth in SEQ ID NOs: 1, 3 or a portion thereof.

46. A method of increasing a level of free glucose in a plant derived glucose containing fermentation starting material, the method comprising the

step of fermenting the plant derived glucose containing fermentation starting material by a cell, said plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the free glucose in the plant.

47. The method of claim 46, wherein said polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.

48. The method of claim 46, wherein said polynucleotide is as set forth in SEQ ID NOs: 1, 3 or a portion thereof.

49. A method of increasing a level of free glucose in a plant, the method comprising the step of overexpressing in the plant a nucleic acid construct including a genomic, complementary or composite polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity, thereby increasing the level of the free glucose in the plant.

50. The method of claim 49, wherein said polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.

51. The method of claim 49, wherein said polynucleotide is as set forth in SEQ ID NOs: 1, 3 or a portion thereof.

52. A method of producing an alcohol, the method comprising the step of fermenting a glucose containing fermentation starting material by a cell overexpressing a nucleic acid construct including a genomic, complementary

or composite polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity, and extracting the alcohol therefrom.

53. The method of claim 52, wherein said cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.

54. The method of claim 53, wherein said prokaryotic cell is *E. coli*.

55. The method of claim 53, wherein said eukaryotic cell is selected from the group consisting of a yeast cell, a fungous cell, a plant cell and an animal cell.

56. The method of claim 52, wherein said polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.

57. The method of claim 52, wherein said polynucleotide is as set forth in SEQ ID NOs: 1, 3 or a portion thereof.

58. A method of producing an alcohol, the method comprising the step of fermenting a plant derived glucose containing fermentation starting material by a cell, said plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity, and extracting the alcohol therefrom.

59. The method of claim 58, wherein said polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.

60. The method of claim 58, wherein said polynucleotide is as set forth in SEQ ID NOs: 1, 3 or a portion thereof.

61. A method of producing an aroma spreading plant, the method comprising the step of overexpressing in the plant a nucleic acid construct including a genomic, complementary or composite polynucleotide being derived from *Aspergillus niger*, encoding a polypeptide having a β -glucosidase catalytic activity, thereby increasing aroma spread from the plant.

62. The method of claim 61, wherein overexpressing said nucleic acid construct is performed in a tissue specific manner.

63. The method of claim 61, wherein overexpressing said nucleic acid construct is limited to at least one tissue selected from the group consisting of flower, fruit, seed, root, stem, pollen and leaves.

64. The method of claim 61, wherein said polypeptide is as set forth in SEQ ID NO: 2 or a portion thereof having said β -glucosidase catalytic activity.

65. The method of claim 61, wherein said polynucleotide is as set forth in SEQ ID NOs: 1, 3 or a portion thereof.

66. An isolated nucleic acid comprising a genomic, complementary or composite polynucleotide encoding a polypeptide having a β -glucosidase catalytic activity and further encoding a signal peptide in frame with said polypeptide.

67. The isolated nucleic acid of claim 66, further encoding an endoplasmic retaining sequence in frame with said polypeptide.

68. A nucleic acid construct comprising the isolated nucleic acid of claim 66.

69. The nucleic acid construct of claim 68, further comprising at least one cis acting control element for regulating expression of said polynucleotide.

70. A host cell comprising the nucleic acid construct of claim 68.

71. The host cell of claim 70, wherein the host cell is of a plant.

72. The host cell of claim 70, wherein the host cell is of a yeast.

73. The host cell of claim 70, wherein the host cell is of an animal.

74. An organism comprising the isolated nucleic acid of claim 66 as an inter or intra chromosomal genetic element.

75. The organism of claim 74, wherein the organism is a plant.

76. A recombinant protein comprising a polypeptide having a β -glucosidase catalytic activity and a signal peptide fused thereto.

77. The recombinant protein of claim 76, further comprising an endoplasmic reticulum retaining peptide fused thereto.

78. A method of producing recombinant β -glucosidase, the method comprising the step of introducing, in an expressible form, a nucleic acid construct into a host cell, said nucleic acid construct including a genomic, complementary or composite polynucleotide encoding a polypeptide having a

β -glucosidase catalytic activity and a signal peptide being in frame with said polypeptide.

79. The method of claim 78, wherein said polynucleotide further encodes an endoplasmic reticulum retaining peptide being in frame with said polypeptide.

80. The method of claim 78, further comprising the step of extracting said polypeptide having said β -glucosidase catalytic activity.

81. The method of claim 78, wherein said host cell is a plant cell.

82. The method of claim 78, wherein said host cell is an animal cell.

83. The method of claim 78, wherein said host cell is a yeast cell.

84. A method of producing a recombinant β -glucosidase overexpressing cell, the method comprising the step of introducing, in an overexpressible form, a nucleic acid construct into a host cell, said nucleic acid construct including a genomic, complementary or composite polynucleotide encoding a polypeptide having a β -glucosidase catalytic activity and further encoding a signal peptide being in frame with said polypeptide.

85. The method of claim 84, wherein said polynucleotide further encodes an endoplasmic reticulum retaining peptide being in frame with said polypeptide.

86. The method of claim 84, wherein said host cell is a plant cell.

87. The method of claim 84, wherein said host cell is a yeast cell.

88. The method of claim 84, wherein said host cell is an animal cell.

89. A method of increasing a level of at least one fermentation substance in a fermentation product, the method comprising the step of fermenting a plant derived glucose containing fermentation starting material by a yeast cell, said plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide encoding a polypeptide having a β -glucosidase catalytic activity and further encoding a signal peptide being in frame with said polypeptide, thereby increasing the level of the at least one fermentation substance in the fermentation product.

90. The method of claim 89, wherein said polynucleotide further encodes an endoplasmic reticulum retaining peptide being in frame with said polypeptide.

91. A method of increasing a level of at least one aroma substance in a plant derived product, the method comprising the step of incubating a glucose containing plant starting material with a yeast cell, said plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide encoding a polypeptide having a β -glucosidase catalytic activity and further encoding a signal peptide being in frame with said polypeptide, thereby increasing the level of the at least one aroma substance in the plant derived product.

92. The method of claim 91, wherein said polynucleotide further encodes an endoplasmic reticulum retaining peptide being in frame with said polypeptide.

93. The method of claim 91, wherein said plant derived product is a fermentation product.

94. A method of increasing a level of free glucose in a glucose containing fermentation starting material, the method comprising the step of fermenting the glucose containing fermentation starting material by a cell overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide encoding a polypeptide having a β -glucosidase catalytic activity and further encoding a signal peptide being in frame with said polypeptide, thereby increasing the level of the free glucose in the glucose containing fermentation starting material.

95. The method of claim 94, wherein said polynucleotide further encodes an endoplasmic reticulum retaining peptide being in frame with said polypeptide.

96. A method of increasing a level of free glucose in a plant derived glucose containing fermentation starting material, the method comprising the step of fermenting the plant derived glucose containing fermentation starting material by a cell, said plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide encoding a polypeptide having a β -glucosidase catalytic activity and further encoding a signal peptide being in frame with said polypeptide, thereby increasing the level of the free glucose in the plant.

97. The method of claim 96, wherein said polynucleotide further encodes an endoplasmic reticulum retaining peptide being in frame with said polypeptide.

98. A method of increasing a level of free glucose in a plant, the method comprising the step of overexpressing in the plant a nucleic acid construct including a genomic, complementary or composite polynucleotide encoding a polypeptide having a β -glucosidase catalytic activity and further encoding a signal peptide being in frame with said polypeptide, thereby increasing the level of the free glucose in the plant.

99. The method of claim 98, wherein said polynucleotide further encodes an endoplasmic reticulum retaining peptide being in frame with said polypeptide.

100. A method of producing an alcohol, the method comprising the step of fermenting a glucose containing fermentation starting material by a cell overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide encoding a polypeptide having a β -glucosidase catalytic activity and further encoding a signal peptide being in frame with said polypeptide, and extracting the alcohol therefrom.

101. The method of claim 100, wherein said polynucleotide further encodes an endoplasmic reticulum retaining peptide being in frame with said polypeptide.

102. A method of producing an alcohol, the method comprising the step of fermenting a plant derived glucose containing fermentation starting material by a cell, said plant overexpressing a nucleic acid construct including a genomic, complementary or composite polynucleotide encoding a polypeptide having a β -glucosidase catalytic activity and further encoding a signal peptide being in frame with said polypeptide, and extracting the alcohol therefrom.

103. The method of claim 102, wherein said polynucleotide further encodes an endoplasmic reticulum retaining peptide being in frame with said polypeptide.

104. A method of producing an aroma spreading plant, the method comprising the step of overexpressing in the plant a nucleic acid construct including a genomic, complementary or composite polynucleotide encoding a polypeptide having a β -glucosidase catalytic activity and further encoding a signal peptide being in frame with said polypeptide, thereby increasing aroma spread from the plant.

105. The method of claim 104, wherein said polynucleotide further encodes an endoplasmic reticulum retaining peptide being in frame with said polypeptide.

106. The method of claim 104, wherein overexpressing said nucleic acid construct is performed in a tissue specific manner.

107. The method of claim 104, wherein overexpressing said nucleic acid construct is limited to at least one tissue selected from the group consisting of flower, fruit, seed, root, stem, pollen and leaves.

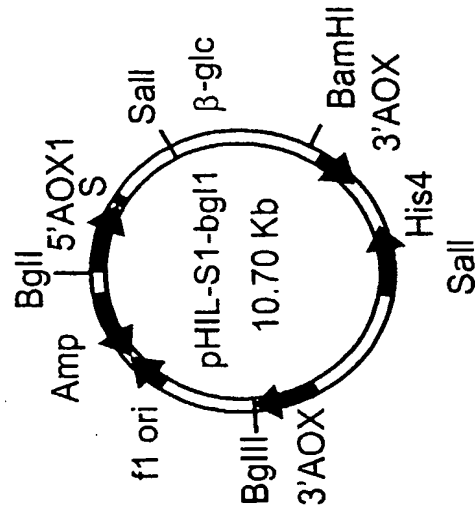


Fig. 1c

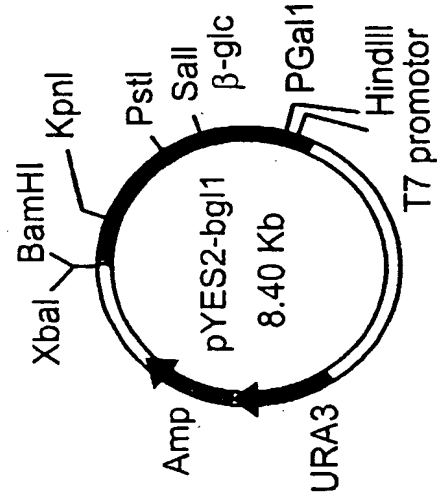


Fig. 1b

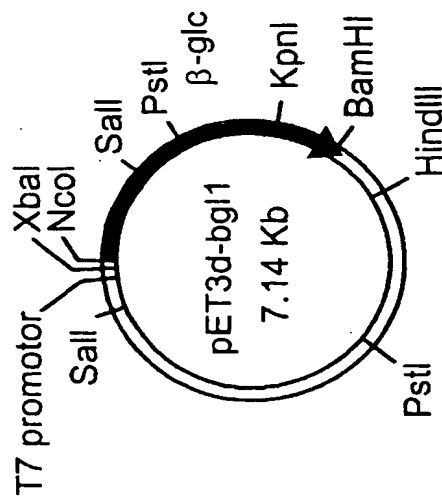


Fig. 1a

2/12

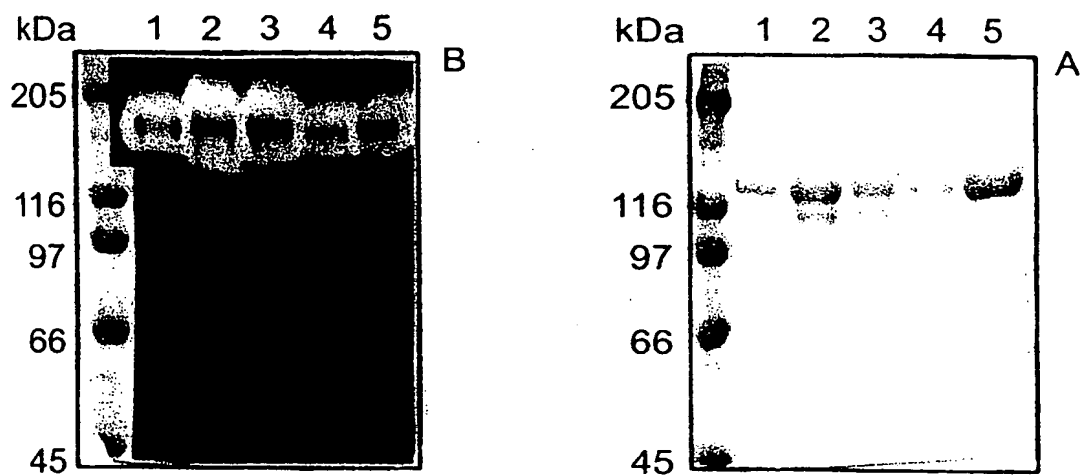


Fig. 2

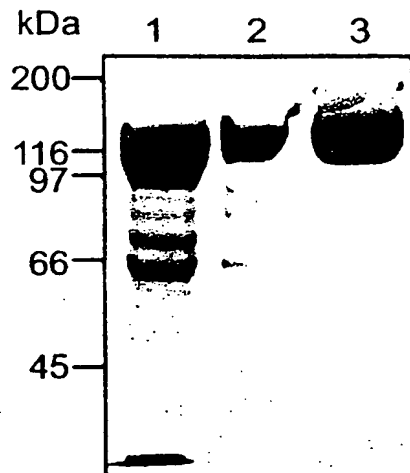


Fig. 3

3/12

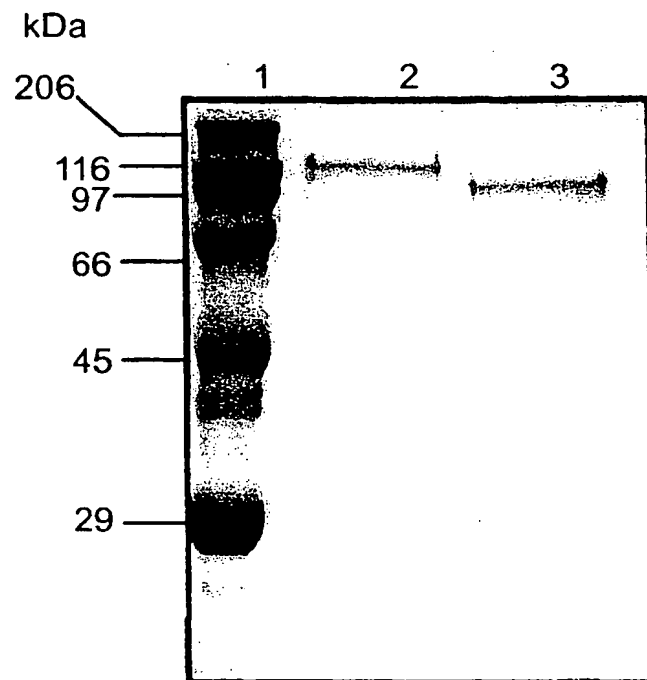


Fig. 4

4/12

1	TCCATTCGCCCCATGCTTAGCGTGTCTTTTCTTTGAACACGCGATGCGGGACTGTGAATTG	60
61	CATGAGTGGGTAGCTTTGCGGAGACAGCTGCACTGGCATAATCATCGTTGGGTTCCCTCA	120
121	ATTCCGATGCCGTGGCGGACGGTCACTTTGTGGCGCTCAAACATTTTAATATGGCCCAGC	180
181	TCCCCCTTCTCTCGCTGTTTTCTGTTCTGTCTCCCTAAACCTCCAGTCTCTCCATTGGA	240
241	CAGGTGTGACAGGTTGCTCACCTGGTTTGTGTTTGTCTCCCTTTGGGCGACCTTGCCAT	300
301	CATGAGGTTCACTTTGATCGAGGCGGTGGCTCTGACTGCCGTCTCGCTGGCCAGCGCTGT	360
	MetArgPheThrLeuIleGluAlaValAlaLeuThrAlaValSerLeuAlaSerAla	
	Signal Peptide	
361	ACGTGCCGTTACTTTGTCTGAGAATTGCAATTGTGCTTAATTAGATTCAATTGTTTGT	420
	Intron#1	
421	TCATCATCGCTGACAATGGTCTTTTCATAGGATGAATTGGCCTACTCCCCACCGTATTAC	480
	AspGluLeuAlaTyrSerProProTyrTyr	
481	CCATCCCCCTTGGGCCAATGGCCAGGGCGACTGGGCGCAGGCATACCAGCGCGCTGTTGAT	540
	ProSerProTrpAlaAsnGlyGlnGlyAspTrpAlaGlnAlaTyrGlnArgAlaValAsp	
541	ATTGTCTCGCAAATGACATTGGATGAGAAGGTCAATCTGACCACAGGAAGTGGGTAGGGC	600
	IleValSerGlnMetThrLeuAspGluLysValAsnLeuThrThrGlyThrGly	
601	TTACATGGCGCAATCTGTATGCTCCGGCTAACAACCTCTACATGGGAATTGGAAGTATGT	660
	Intron#2 TrpGluLeuGluLeuCys	
661	GTTGGTCAGACTGGCGGTGTTCCCCGGTAGGTTTGAAAATATTGTCGAGACAGGGGACAT	720
	ValGlyGlnThrGlyGlyValProArg Intron#3	
721	TATTGATTACGGTGACAGATTGGGAGTTCGGGAATGTGTTTACAGGATAGCCCTCTGG	780
	LeuGlyValProGlyMetCysLeuGlnAspSerProLeuG	
781	GCGTTCGCGACTGTAAGCCATCTGCTGTGTTAGGCTTCGATGCTCTTACTGACACGGCG	840
	lyValArgAspS Intron#4	
841	CAGCCGACTACAACCTCTGCTTTCCCTGCCGGCATGAACGTGGCTGCAACCTGGGACAAGA	900
	erAspTyrAsnSerAlaPheProAlaGlyMetAsnValAlaAlaThrTrpAspLysA	
901	ATCTGGCATACTTCGCGGCAAGGCTATGGGTGAGGAATTTAGTGACAAGGGTGCCGATA	960
	snLeuAlaTyrLeuArgGlyLysAlaMetGlyGlnGluPheSerAspLysGlyAlaAspI	
961	TCCAATTGGGTCCAGCTGCCGGCCCTCTCGGTAGAAGTCCCGACGGTGGTTCGTAACCTGGG	1020
	leGlnLeuGlyProAlaAlaGlyProLeuGlyArgSerProAspGlyGlyArgAsnTrpG	
1021	AGGGCTTCTCCCAGACCCTGCCCTAAGTGGTGTGCTCTTTGCCGAGACCATCAAGGGTA	1080
	luGlyPheSerProAspProAlaLeuSerGlyValLeuPheAlaGluThrIleLysGlyI	
1081	TCCAAGATGCTGGTGTGGTTGCGACGGCTAAGCACTACATTGCTTACGAGCAAGAGCATT	1140
	leGlnAspAlaGlyValValAlaThrAlaLysHisTyrIleAlaTyrGluGlnGluHisP	
1141	TCCGTGAGGCGCTGAAGCCCAAGGTTTTGGATTTAATATTTCCGAGAGTGGAAAGTGCGA	1200
	heArgGlnAlaProGluAlaGlnGlyPheGlyPheAsnIleSerGluSerGlySerAlaA	
1201	ACCTCGACGATAAGACTATGCACGAGCTGTACCTCTGGCCCTTCGCGGATGCCATCCGTG	1260
	snLeuAspAspLysThrMetHisGluLeuTyrLeuTrpProPheAlaAspAlaIleArgA	
1261	CAGGTGCTGGCGCTGTGATGTGCTCCTACAACCAGATCAACAACAGTTATGGCTGCCAGA	1320
	laGlyAlaGlyAlaValMetCysSerTyrAsnGlnIleAsnAsnSerTyrGlyCysGlnA	
1321	ACAGCTACACTCTGAACAAGCTGCTCAAGGCCGAGCTGGGCTTCCAGGGCTTTGTCTATGA	1380
	snSerTyrThrLeuAsnLysLeuLeuLysAlaGluLeuGlyPheGlnGlyPheValMetS	

Fig. 5a

5/12

1381 GTGATTGGGCTGCTCACCATGCTGGTGTGAGTGGTGCTTTGGCAGGATTGGATATGTCTA 1440
 erAspTrpAlaAlaHisHisAlaGlyValSerGlyAlaLeuAlaGlyLeuAspMetSerM
 1441 TGCCAGGAGACGTCGACTACGACAGTGGTACGTCTTACTGGGGTACAACTTGACCATTA 1500
 etProGlyAspValAspTyrAspSerGlyThrSerTyrTrpGlyThrAsnLeuThrIleS
 1501 GCGTGCTCAACGGAACGGTGCCCCAATGGCGTGTGATGACATGGCTGTCCGCATCATGG 1560
 erValLeuAsnGlyThrValProGlnTrpArgValAspAspMetAlaValArgIleMetA
 1561 CCGCCTACTACAAGGTGCGCCGTGACCGTCTGTGGACTCCTCCCACTTCAGCTCATGGA 1620
 laAlaTyrTyrLysValGlyArgAspArgLeuTrpThrProProAsnPheSerSerTrpT
 1621 CCAGAGATGAATACGGCTACAAGTACTACTACGTGTGCGAGGGACCGTACGAGAAGGTCA 1680
 hrArgAspGluTyrGlyTyrLysTyrTyrTyrValSerGluGlyProTyrGluLysValA
 1681 ACCAGTACGTGAATGTGCAACGCAACCACAGCGAAGTATTGCGCCGATTGGAGCGGACA 1740
 snGlnTyrValAsnValGlnArgAsnHisSerGluLeuIleArgArgIleGlyAlaAspS
 1741 GCACGGTGCTCCTCAAGAACGACGGCGCTCTGCCTTTGACTGGTAAGGAGCGCCTGGTCCG 1800
 erThrValLeuLeuLysAsnAspGlyAlaLeuProLeuThrGlyLysGluArgLeuValA
 1801 CGCTTATCGGAGAAGATGCGGGCTCCAACCTTATGGTGCCAACGGCTGCAGTGACCGTG 1860
 laLeuIleGlyGluAspAlaGlySerAsnProTyrGlyAlaAsnGlyCysSerAspArgG
 1861 GATCGGACAATGGAACATTGGCGATGGGCTGGGGAAGTGGTACTGCCAACTTCCCATACC 1920
 lyCysAspAsnGlyThrLeuAlaMetGlyTrpGlySerGlyThrAlaAsnPheProTyrL
 1921 TGGTGACCCCGAGCAGGCCATCTCAAACGAGGTGCTTAAGCACAAGAATGGTGTATTCA 1980
 euValThrProGluGlnAlaIleSerAsnGluValLeuLysHisLysAsnGlyValPheT
 1981 CCGCCACCGATAACTGGGCTATCGATCAGATTGAGGCGCTTGCTAAGACCGCCAGGTAAG 2040
 hrAlaThrAspAsnTrpAlaIleAspGlnIleGluAlaLeuAlaLysThrAlaArg
 2041 AAGATCCCGGATTCTTTTCCTTCTTGTGCAATGGATGCTGACAACATGCTAGTGTCTCTC 2100
 Intron#5 ValSerL
 2101 TTGTCTTTGTCAACGCGACTCTGGTGAGGGTTACATCAATGTGGACGGAACCTGGGTG 2160
 euValPheValAsnAlaAspSerGlyGluGlyTyrIleAsnValAspGlyAsnLeuGlyA
 2161 ACCGCAGGAACCTGACCCTGTGGAGGAACCGCGATAATGTGATCAAGGCTGCTGCTAGCA 2220
 spArgArgAsnLeuThrLeuTrpArgAsnArgAspAsnValIleLysAlaAlaAlaSerA
 2281 ACCACAACCCCAATGTTACCGCTATCCTCTGGGGTGGTTGCCCCGTCAGGAGTCTGGCA 2340
 snHisAsnProAsnValThrAlaIleLeuTrpGlyGlyLeuProGlyGlnGluSerGlyA
 2341 ACTCTCTTGCCGACGTCCTCTATGGCCGTGTCAACCCCGGTGCCAAGTCGCCCTTTACCT 2400
 snSerLeuAlaAspValLeuTyrGlyArgValAsnProGlyAlaLysSerProPheThrT
 2401 GGGGCAAGACTCGTGAGGCCTACCAAGACTACTTGGTCACCGAGCCCAACAACGGCAACG 2460
 rpGlyLysThrArgGluAlaTyrGlnAspTyrLeuValThrGluProAsnAsnGlyAsnG
 2461 GAGCCCTCAGGAAGACTTTGTGCGAGGGCGTCTTCATTGACTACCGTGGATTGACAAGC 2520
 lyAlaProGlnGluAspPheValGluGlyValPheIleAspTyrArgGlyPheAspLysA
 2521 GCAACGAGACCCCGATCTACGAGTTCGGCTATGGTCTGAGCTACGCCACTTTCAACTACT 2580
 rgAsnGluThrProIleTyrGluPheGlyTyrGlyLeuSerTyrAlaThrPheAsnTyrS
 2581 CGAACCTTGAGGTGCAGGTGCTGAGCGCCCTGCATACGAGCCTGCTTCGGGTGAGACCG 2640
 erAsnLeuGluValGlnValLeuSerAlaProAlaTyrGluProAlaSerGlyGluThrG
 2701 TGCAGAGAATTACCAAGTTCATCTACCCCTGGCTCAACGGTACCGATCTCGAGGCATCTT 2760
 luGlnArgIleThrLysPheIleTyrProTrpLeuAsnGlyThrAspLeuGluAlaSerS
 2761 CCGGGGATGCTAGCTACGGGCAGGACTCCTCCGACTATCTTCCGAGGGAGCCACCGATG 2820
 erGlyAspAlaSerTyrGlyGlnAspSerSerAspTyrLeuProGluGlyAlaThrAspG

Fig. 5a (continued)

6/12

2821 GCTCTGCGCAACCGATCCTGCCTGCCGGTGGCGGTCTGGCGGCAACCCCTCGCCTGTACG 2880
lySerAlaGlnProIleLeuProAlaGlyGlyGlyProGlyGlyAsnProArgLeuTyrA

2881 ACGAGCTCATCCGCGTGTCACTGACCATCAAGAACACCGGCAAGGTTGCTGGTGATGAAG 2940
spGluLeuIleArgValSerValThrIleLysAsnThrGlyLysValAlaGlyAspGluV

2941 TTCCCCAACTGGTAAGTAAACATGAGGTCCGAAAGAGGTTGAACAAAGCTAATCAGTCGC 3000
alProGlnLeu Intron#6

3001 AGTATGTTTCCCTTGGCGGTCCCAATGAGCCCAAGATCGTGCTGCGTCAATTCGAGCGCA 3060
TyrValSerLeuGlyGlyProAsnGluProLysIleValLeuArgGlnPheGluArgI

3061 TCACGCTGCAGCCGTGGGAGGAGACGAAGTGGAGCAGCACTCTGACGCGCCGTGACCTTG 3120
leThrLeuGlnProSerGluGluThrLysTrpSerThrThrLeuThrArgArgAspLeuA

3121 CAAACTGGAATGTTGAGAAGCAGGACTGGGAGATTACGTCGTATCCCAAGATGGTGTGTTG 3180
laAsnTrpAsnValGluLysGlnAspTrpGluIleThrSerTyrProLysMetValPheV

3181 TCGGAAGCTCCTCGCGGAAGCTGCCGCTCCGGGCGTCTCTGCCTACTGTTCACTAAATAG 3240
alGlySerSerSerArgLysLeuProLeuArgAlaSerLeuProThrValHis***

3241 CTCTCAAATGGTATACCATGATGGCCGTGGTATATGAATTAATGATTTATGCCAACAGCA 3300

3301 AGACCACTGTAGATGTAGATGTAGAATGAGTATTGCGTAGTAGCGTGTAGATGATGATAC 3360

3361 AAGCGATCCGACACATGGTAGGAAGAGTGGCGCTAGTTGGGGCGGAAACCAAGCGACGTC 3420

3421 ATCCGCTGCGACTTCGCCAGTCTTTCTTCTTTCTCTTTCAGCCTTCTTCTCCGCTTA 3480

3481 ATCCAGCAACCATTGCCAATTGCCTCTACAACAATAATTGCCATAATACTCTACTCCTA 3540

3541 TTCAATATATACACCACAATCTCGACATAATCACACAAGCCTGAACACACGAGCAACCAT 3600

3601 GCCCTCTCCCGATCCTCCAGCCCCAGCGATACGACCCCTTCCAACCACCCATAACAGCGCT 3660

3661 CCTCATCTACCCAGCGACCCTAATCGTGGGATCACTCTTCTCCGTCCTCTCTCCACCGC 3720

3721 ACAAGGCACACGCGACGACGGCTCCAGCACCCCTCCACCCACACGTCGAGCCCCCTAGCCCC 3780

3781 GTCCATCGCGTCAGACCTCAACCTCTCCTTTCTCCGCCGCGCCCCGTCAACTACTTCGC 3840

3841 TCGCAAAGACAACATCTTCAATCTATATTTCGTCAAAGTCGGC 3885

Fig. 5a (continued)



Fig. 5b

7/12

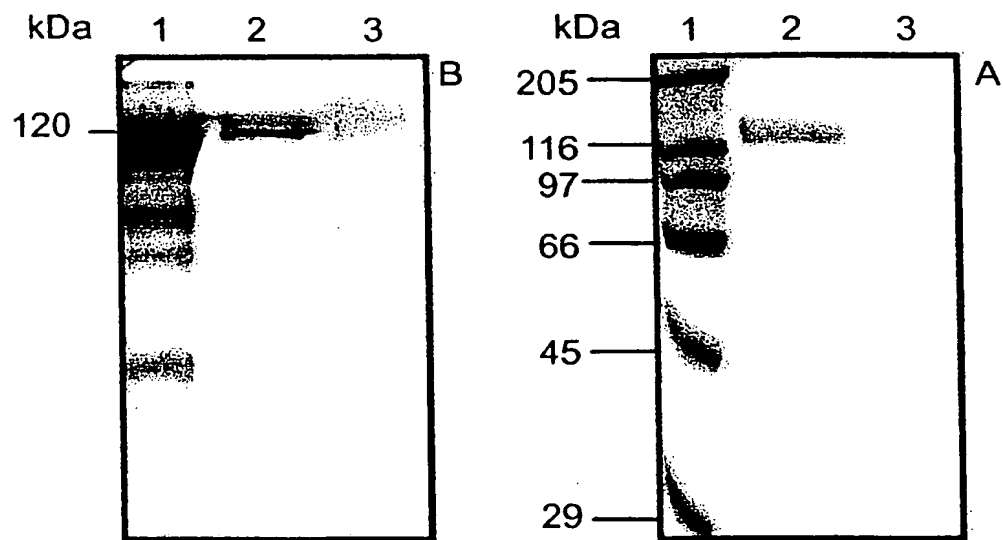


Fig. 6

8/12

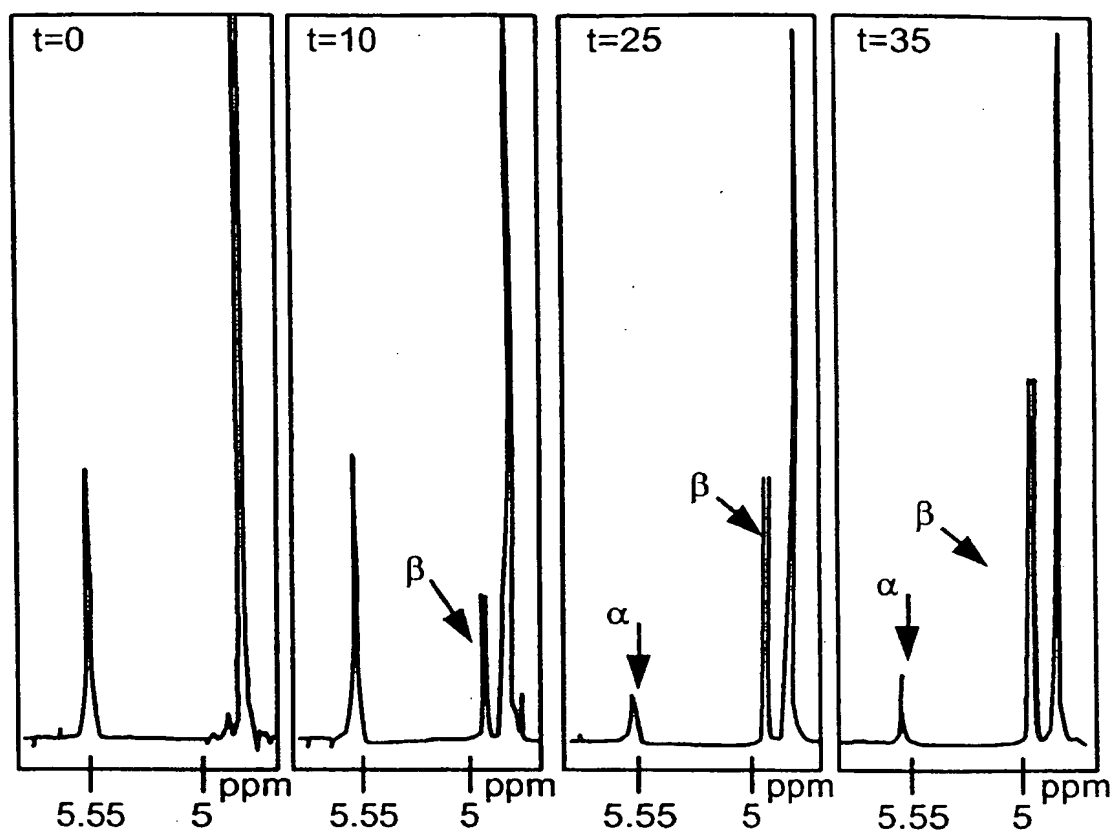


Fig. 7

9/12

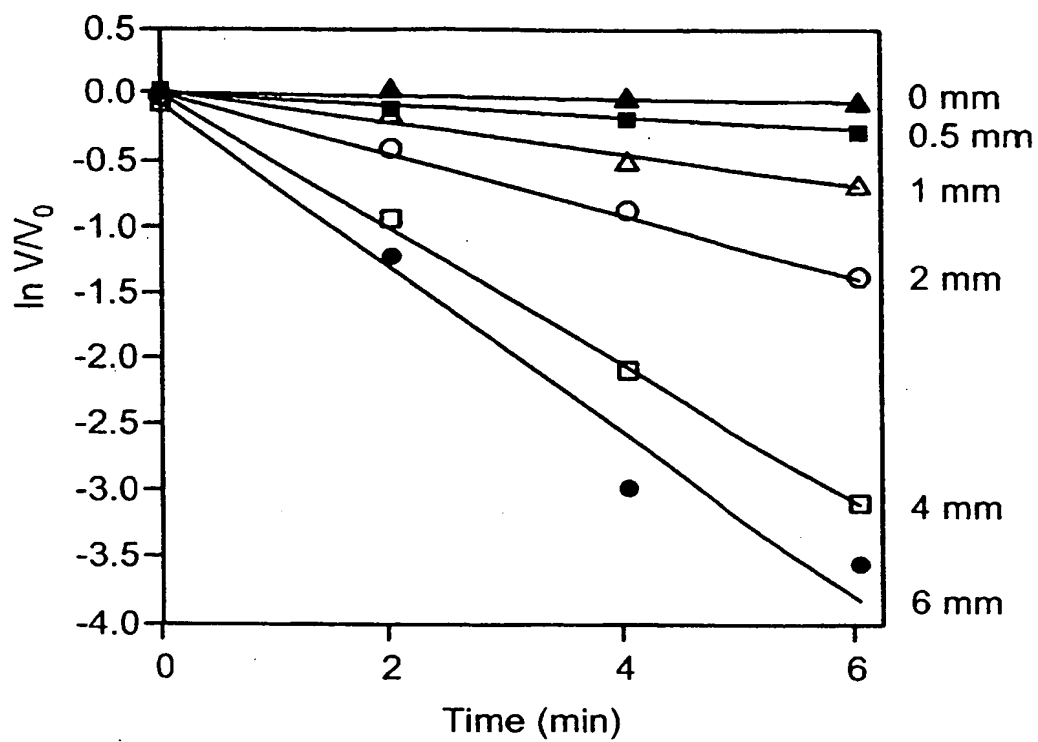


Fig. 8

10/12

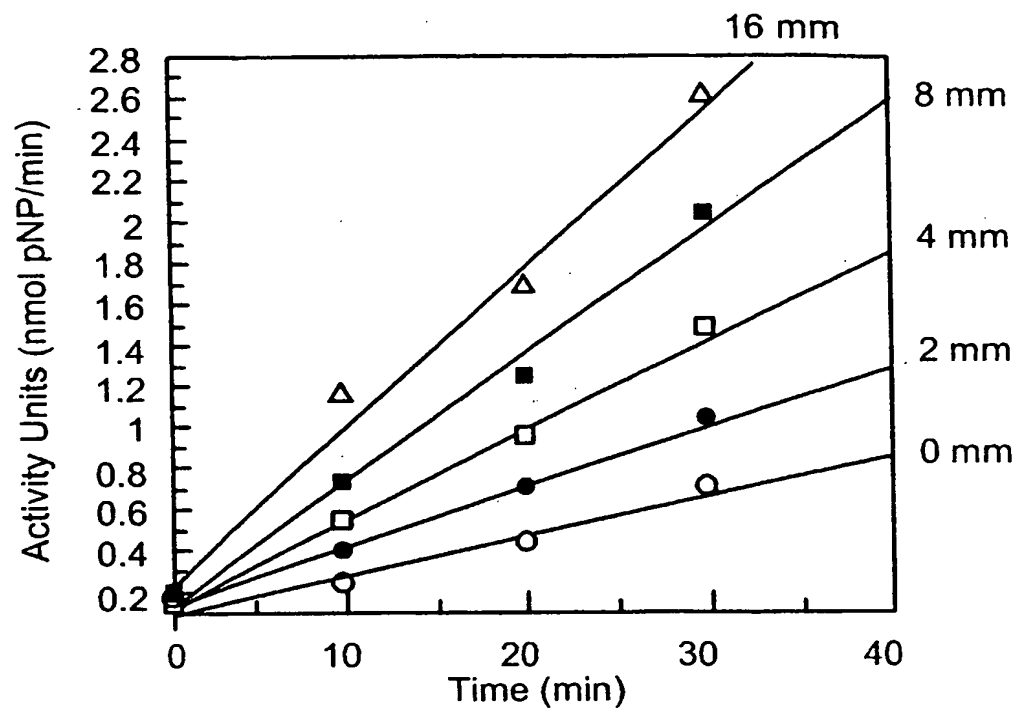


Fig. 9

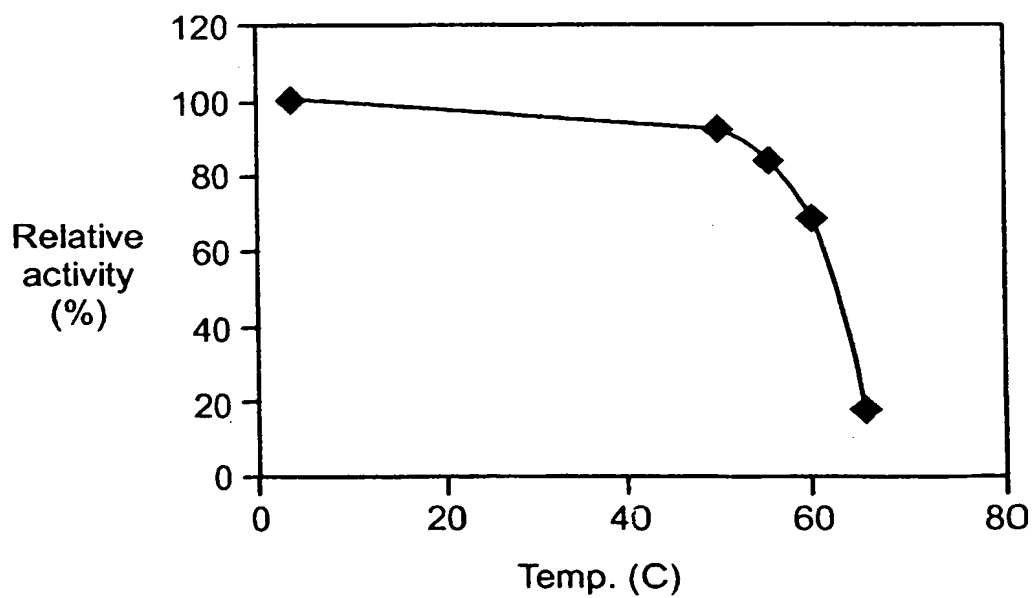


Fig. 10

SUBSTITUTE SHEET (RULE 26)

11/12

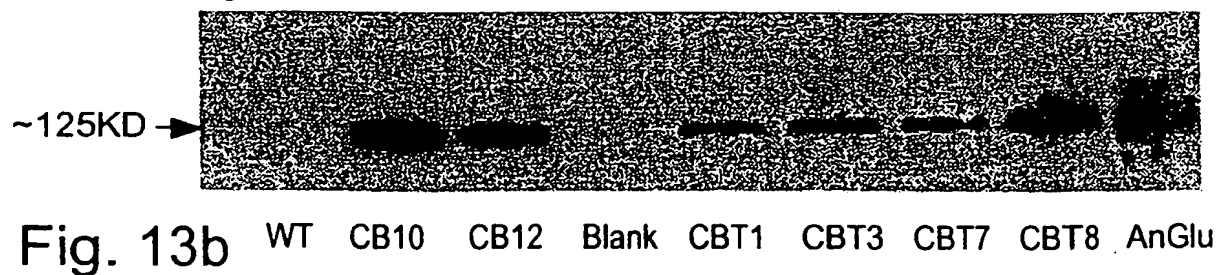
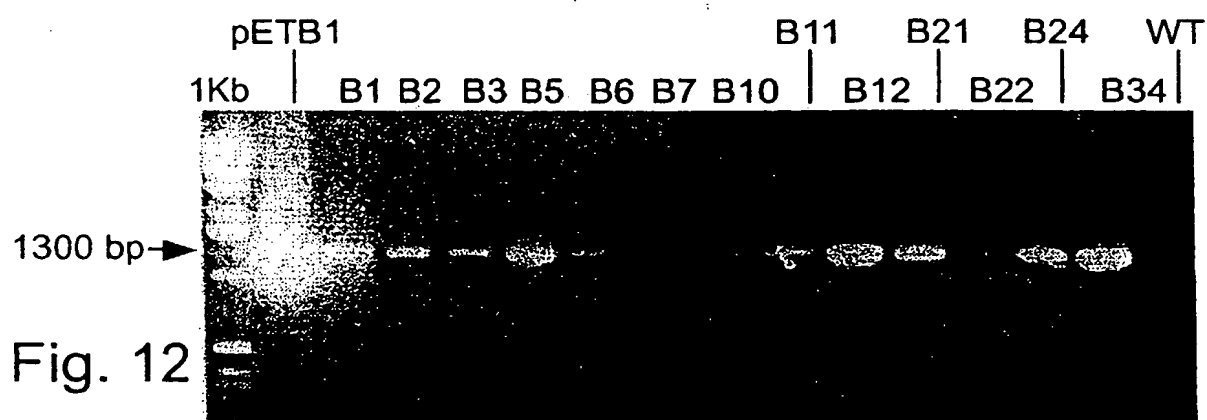
Fig. 11a

35S Ω + bg11

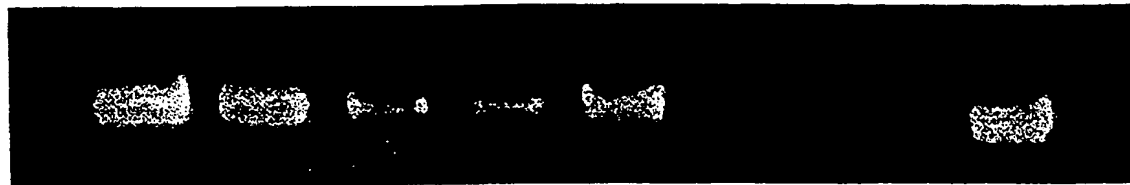
Fig. 11b

35S Ω + Ce11 + bg11

Fig. 11c

35S Ω + Ce11 + bg11 + HDEL

12/12



WT CB10 CB11 CBT3 CBT8 CBT15 B1 B34 An Glu

Fig. 14

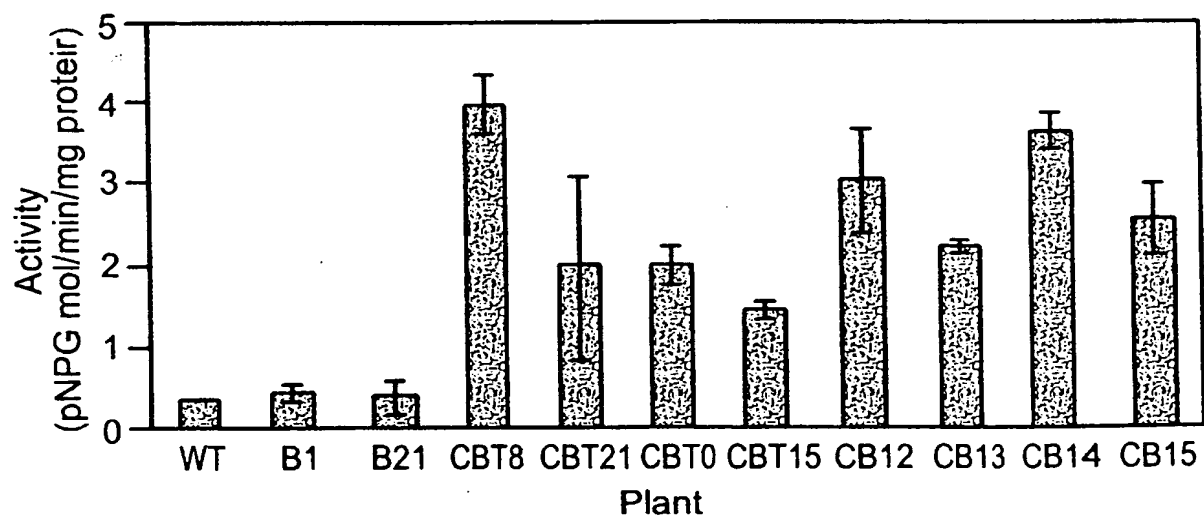


Fig. 15

1
SEQUENCE LISTING

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2583
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

ATGAGGTTCA CTTTGATCGA GGCGGTGGCT CTGACTGCCG TCTCGCTGGC 50
CAGCGCTGAT GAATTGGCCT ACTCCCCACC GTATTACCCA TCCCCTTGGG 100
CCAATGGCCA GGGCGACTGG GCGCAGGCAT ACCAGCGCGC TGTTGATATT 150
GTCTCGCAAA TGACATTGGA TGAGAAGGTC AATCTGACCA CAGGAAGTGG 200
ATGGGAATTG GAACTATGTG TTGGTCAGAC TGGCGGTGTT CCCCATTGG 250
GAGTTCCGGG AATGTGTTTA CAGGATAGCC CTCTGGGCGT TCGCGACTCC 300
GACTACAACCT CTGCTTTCCC TGCCGGCATG AACGTGGCTG CGACCTGGGA 350
CAAGAATCTG GCATACCTTC GCGGCAAGGC TATGGGTCAG GAATTTAGTG 400
ACAAGGGTGC CGATATCCAA TTGGGTCCAG CTGCCGGCCC TCTCGGTAGA 450
AGTCCCGACG GTGGTCGTAA CTGGGAGGGC TTCTCCCCAG ACCCTGCCCT 500
AAGTGGTGTG CTCTTTGCCG AGACCATCAA GGGTATCCAA GATGCTGGTG 550
TGTTTGGCAG GGCTAAGCAC TACATTGCTT ACGAGCAAGA GCATTTCCGT 600
CAGGCGCTG AAGCCCAAGG TTTTGGATT AATATTCCG AGAGTGGAA 650
TGCGAACCTC GATGATAAGA CTATGCACGA GCTGTACCTC TGGCCCTTCG 700
CGGATGCCAT CCGTGACAGT GCTGGCGCTG TGATGTGCTC CTACAACCAG 750
ATCAACAACA GTTATGGCTG CCAGAACAGC TACTCTCTGA ACAAGCTGCT 800
CAAGGCCGAG CTGGGCTTCC AGGGCTTTGT CATGAGTGAT TGGGCTGCTC 850
ACCATGCTGG TGTGAGTGGT GCTTTGGCAG GATTGGATAT GTCTATGCCA 900
GGAGACGTCG ACTACGACAG TGGTACGTCT TACTGGGGTA CAAACTTGAC 950
CATTAGCGTG CTCAACGGAA CGGTGCCCCA ATGGCGTGTG GATGACATGG 1000
CTGTCCGCAT CATGGCCGCC TACTACAAGG TCGGCCGTGA CCGTCTGTGG 1050
ACTCCTCCCA ACTTCAGCTC ATGGACCAGA GATGAATACG GCTACAAGTA 1100
CTACTACGTG TCGGAGGGAC CGTACGAGAA GGTCAACGAG TACGTGAATG 1150
TGCAACGCAA CCACAGCGAA CTGATTCGCC GCATTGGAGC GGACAGCACG 1200
GTGCTCCTCA AGAAGCAGCG CGCTCTGCCT TTGACTGGTA AGGAGCGCCT 1250
GGTCGCGCTT ATCGAGAGAG ATGCGGGGCT CAACCCCTAT GGTGCCAACC 1300
GCTGCAGTGA CCGTGGATGC GACAATGGAA CATGGCGAT GGGCTGGGGA 1350
AGTGGTACTG CCAACTTCCC ATACCTGGTG ACCCCCGAGC AGGCCATCTC 1400
AAACGAGGTG CTTAAGCACA AGAATGGTGT ATTCAACGCC ACCGATAACT 1450
GGGCTATCGA TCAAATTGAG GCGCTTGCTA AGACCGCCAG TGTCTCTCTT 1500
GTCTTTGTCA ACGCCGACTC TGGTGAGGGT TACATCAATG TGGACGGAAA 1550
CCTGGGTGAC CGCAGGAACC TGACCCTGTG GAGGAACGGC GATAATGTGA 1600
TCAAGGCTGC TGCTAGCAAC TGCAACAACA CAATCGTTGT CATTCACTCT 1650
TTCGGACGAC TCTTGGTTAA CGAGTGGTAC GACRACCCCA ATGTTACCGC 1700
TATCTCTTGG GTGGGTTTGC CCGGTCAGGA GTCTGGCAAC TCTCTTGCCG 1750
ACGTCTCTTA TGGCGGTGTC AACCCTGGTG CCAAGTCGCC CTTTACCTGG 1800
GGCAAGACTC GTGAGGCTTA CCAAGACTAC TTGGTCACCG AGCCCAACAA 1850
CGGCAACGGA GCCCTCAGG AAGACTTTGT CGAGGGCGTC TTCAATTGACT 1900
ACCGTGGATT TGACAAGCGC AACGAGACCC CGATCTACGA GTTCGGCTAT 1950
GGCTGAGCT ACACCACTTT CAACTACTCG AACCTTGAGG TGCAGGTGCT 2000
GAGCGCCCT GCATACGAGC CTGCTTCGGG TGAGACCGAG GCAGCGCCAA 2050
CCTTCGGAGA GGTGGAAAT GCGTCGGATT ACCTTACCC CAGCGGATTG 2100
CTGAGAATTA CCAAGTTCAT CTACCCCTGG CTCACCGGTA CCGATCTCGA 2150
GGCATCTTCC GGGGATGCTA GCTACGGGCA GGACTCTCTC GACTATCTTC 2200
CCGAGGGAGC CACCGATGGC TCTGCGCAAC CGATCCTGCC TGCCGGTGGC 2250
GGTCTGGCG GCAACCTCG CCTGTACGAC GAGCTCATCC GCGTGTCACT 2300
GACCATCAAG AACACCGGCA AGGTTGCTGG TGATGAAGTT CCCCAGTGT 2350
ATGTTTCCCT TGGCGGTCCC AATGAGCCCA AGATCGTCTG GCGTCAATTC 2400
GAGCGCATCA CGCTGCAGCC GTCGGAGGAG ACGAAGTGGA GCACGACTCT 2450
GACGCGCCGT GACCTTGCAA ACTGGAATGT TGAGAAGCAG GACTGGGAGA 2500
TTACGTCGTA TCCAAGATG GTGTTGTGCG GAAGCTCCTC GCGGAAGCTG 2550
CCGCTCCGGG CGTCTCTGCC TACTGTTTAC TAA 2583

```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 860
(B) TYPE: amino acid
(C) STRAN: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Met Arg Phe Thr Leu Ile Glu Ala Val Ala Leu Thr Ala Val Ser
5 10 15
Leu Ala Ser Ala Asp Glu Leu Ala Tyr Ser Pro Pro Tyr Tyr Pro
20 25 30
Ser Pro Trp Ala Asn Gly Gln Gly Asp Trp Ala Gln Ala Tyr Gln
35 40 45
Arg Ala Val Asp Ile Val Ser Gln Met Thr Leu Asp Glu Lys Val
50 55 60
Asn Leu Thr Thr Gly Thr Gly Trp Glu Leu Glu Leu Cys Val Gly
65 70 75
Gln Thr Gly Gly Val Pro Arg Leu Gly Val Pro Gly Met Cys Leu

```

80	85	90
Gln Asp Ser Pro	Leu Gly Val Arg Asp	Ser Asp Tyr Asn Ser
95	100	105
Phe Pro Ala Gly	Met Asn Val Ala Ala	Thr Trp Asp Lys Asn
110	115	120
Ala Tyr Leu Arg	Gly Lys Ala Met Gly	Gln Glu Phe Ser Asp
125	130	135
Gly Ala Asp Ile	Gln Leu Gly Pro Ala	Ala Gly Pro Leu Gly
140	145	150
Ser Pro Asp Gly	Gly Arg Asn Trp Glu	Gly Phe Ser Pro Asp
155	160	165
Ala Leu Ser Gly	Val Leu Phe Ala Glu	Thr Ile Lys Gly Ile
170	175	180
Asp Ala Gly Val	Val Ala Thr Ala Lys	His Tyr Ile Ala Tyr
185	190	195
Gln Glu His Phe	Arg Gln Ala Pro Glu	Ala Gln Gly Phe Gly
200	205	210
Asn Ile Ser Glu	Ser Gly Ser Ala Asn	Leu Asp Asp Lys Thr
215	220	225
His Glu Leu Tyr	Leu Trp Pro Phe Ala	Asp Ala Ile Arg Ala
230	235	240
Ala Gly Ala Val	Met Cys Ser Tyr Asn	Gln Ile Asn Asn Ser
245	250	255
Gly Cys Gln Asn	Ser Tyr Thr Leu Asn	Lys Leu Leu Lys Ala
260	265	270
Leu Gly Phe Gln	Gly Phe Val Met Ser	Asp Trp Ala Ala His
275	280	285
Ala Gly Val Ser	Gly Ala Leu Ala Gly	Leu Asp Met Ser Met
290	295	300
Gly Asp Val Asp	Tyr Asp Ser Gly Thr	Ser Tyr Trp Gly Thr
305	310	315
Leu Thr Ile Ser	Val Leu Asn Gly Thr	Val Pro Gln Trp Arg
320	325	330
Asp Asp Met Ala	Val Arg Ile Met Ala	Ala Tyr Tyr Lys Val
335	340	345
Arg Asp Arg Leu	Trp Thr Pro Pro Asn	Phe Ser Ser Trp Thr
350	355	360
Asp Glu Tyr Gly	Tyr Lys Tyr Tyr Tyr	Val Ser Glu Gly Pro
365	370	375
Glu Lys Val Asn	Gln Tyr Val Asn Val	Gln Arg Asn His Ser
380	385	390
Leu Ile Arg Arg	Ile Gly Ala Asp Ser	Thr Val Leu Leu Lys
395	400	405
Asp Gly Ala Leu	Pro Leu Thr Gly Lys	Glu Arg Leu Val Ala
410	415	420
Ile Gly Glu Asp	Ala Gly Ser Asn Pro	Tyr Gly Ala Asn Gly
425	430	435
Ser Asp Arg Gly	Cys Asp Asn Gly Thr	Leu Ala Met Gly Trp
440	445	450
Ser Gly Thr Ala	Asn Phe Pro Tyr Leu	Val Thr Pro Glu Gln
455	460	465
Ile Ser Asn Glu	Val Leu Lys His Lys	Asn Gly Val Phe Thr
470	475	480
Thr Asp Asn Trp	Ala Ile Asp Gln Ile	Glu Ala Leu Ala Lys
485	490	495
Ala Ser Val Ser	Leu Val Phe Val Asn	Ala Asp Ser Gly Glu
500	505	510
Tyr Ile Asn Val	Asp Gly Asn Leu Gly	Asp Arg Arg Asn Leu
515	520	525
Leu Trp Arg Asn	Gly Asp Asn Val Ile	Lys Ala Ala Ala Ser
530	535	540
Cys Asn Asn Thr	Ile Val Val Ile His	Ser Val Gly Pro Val
545	550	555
Val Asn Glu Trp	Tyr Asp Asn Pro Asn	Val Thr Ala Ile Leu
560	565	570
Gly Gly Leu Pro	Gly Gln Glu Ser Gly	Asn Ser Leu Ala Asp
575	580	585
Leu Tyr Gly Arg	Val Asn Pro Gly Ala	Lys Ser Pro Phe Thr
590	595	600
Gly Lys Thr Arg	Glu Ala Tyr Gln Asp	Tyr Leu Val Thr Glu
605	610	615
Asn Asn Gly Asn	Gly Ala Pro Gln Glu	Asp Phe Val Glu Gly
620	625	630
Phe Ile Asp Tyr	Arg Gly Phe Asp Lys	Arg Asn Glu Thr Pro
635	640	645
Tyr Glu Phe Gly	Tyr Gly Leu Ser Tyr	Thr Thr Phe Asn Tyr
650	655	660
Asn Leu Glu Val	Gln Val Leu Ser Ala	Pro Ala Tyr Glu Pro
665	670	675
Ser Gly Glu Thr	Glu Ala Ala Pro Thr	Phe Gly Glu Val Gly
		Asn

3

Ala Ser Asp Tyr	680	Leu Tyr Pro Ser Gly	685	Leu Leu Arg Ile Thr	690
	695		700		705
Phe Ile Tyr Pro	710	Trp Leu Asn Gly Thr	715	Asp Leu Glu Ala Ser	720
	725		730		735
Gly Asp Ala Ser	740	Tyr Gly Gln Asp Ser	745	Ser Asp Tyr Leu Pro	750
	755		760		765
Gly Ala Thr Asp	770	Gly Ser Ala Gln Pro	775	Ile Leu Pro Ala Gly	780
	785		790		795
Gly Pro Gly Gly	800	Asn Pro Arg Leu Tyr	805	Asp Glu Leu Ile Arg	810
	815		820		825
Ser Val Thr Ile	830	Lys Asn Thr Gly Lys	835	Val Ala Gly Asp Glu	840
	845		850		855
Pro Gln Leu Tyr	860	Val Ser Leu Gly Gly		Pro Asn Glu Pro Lys	
Val Leu Arg Gln		Phe Glu Arg Ile Thr		Leu Gln Pro Ser Glu	
Thr Lys Trp Ser		Thr Thr Leu Thr Arg		Arg Asp Leu Ala Asn	
Asn Val Glu Lys		Gln Asp Trp Glu Ile		Thr Ser Tyr Pro Lys	
Val Phe Val Gly		Ser Ser Ser Arg Lys		Leu Pro Leu Arg Ala	
Leu Pro Thr Val		His			

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3885
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TCCATTCGCC CATGCTTAGC GTGTCTTTTC TTTGAACACT GCATGCGGGA 50
 CTGTGAATTG CATGAGTGGG TAGCTTTGCG GAGACAGCTG CACTGGCATA 100
 CATCATCGTT GGGTTCCTCA ATTGCGATGC CGTGGCGGAC GGTCACCTTTG 150
 TGGCGCTCAA ACTATTAAAT ATGGCCAGC TCCCTTTTCT CTCGCTGTTT 200
 TCGTTTCTGT CCTCCCTAAA CCTCCAGTCT CTCCATTGGA CAGGTGTTGC 250
 ACGTTGCTC ACCTGGTTTG TTTGCTCCC CTTTGGGCG ACCTTGCCAT 300
 CATGAGGTTT ACTTTGATCG AGGCGGTGGC TCTGACTGCC GTCTCGCTGG 350
 CCAGCGCTGT ACGTGCCGT ACTTTGCTCT GAGAATTGCA ATTGTGCTTA 400
 ATTAGATTCA TTTGTTTGT TCATCATCGC TGACAATGGT CTTTTCATAG 450
 GATGAATTGG CCTACTCCCC ACCGTATTAC CCATCCCCCT GGGCCAATGG 500
 CCAGGGCGAC TGGGCGCAGG CATACCAGCG CGCTGTTGAT ATTGTCTCGC 550
 AATGACATT GGATGAGAAG GTCAATCTGA CCACAGGAAC TGGGTAGGGC 600
 TTACATGGCG CAATCTGTAT GCTCCGGCTA ACAACTTCTA CATGGGAATT 650
 GGAATATGT GTTGCTCAGA CTGGCGGTGT TCCCCGGTAG GTTTGAAAAT 700
 ATTGTCGAGA CAGGGGACAT TATTGATTAA CGGTGACAGA TTGGGAGTTC 750
 CGGGAATGTG TTTACAGGAT AGCCCTCTGG GCGTTCGCGA CTGTAAGCCA 800
 TCTGCTGTTG TTAGGCTTCG ATGCTCTTAC TGACACGGCG CAGCCGACTA 850
 CAACTCTGCT TTCCCTGCCG GCATGAACGT GGCTGCAACC TGGGACAAGA 900
 ATCTGGCATA CCTTCGCGGC AAGGCTATGG GTCAGGAATT TAGTGACAAG 950
 GGTGCCGATA TCCAAATTGGG TCCAGCTGCC GGCCTCTCG GTAGAAGTCC 1000
 CGACGGTGGT CGTAACTGGG AGGGCTTCTC CCCAGACCCT GCCCTAAGTG 1050
 GTGTGCTCTT TGCCGAGACC ATCAAGGGTA TCCAAGATGC TGGTGTGGTT 1100
 GCGACGGCTA AGCACTACAT TGCTTACGAG CAAGAGCATT TCCGTCAGGC 1150
 GCCTGAAGCC CAAGGTTTGT GATTAAATAT TTCCGAGAGT GGAAGTGCGA 1200
 ACCTCGATGA TAAGACTATG CACGAGCTGT ACCTCTGGCC CTTTCGCGAT 1250
 GCCATCCGTG CAGGTGCTGG CGCTGTGATG TGCTCCTACA ACCAGATCAA 1300
 CAACAGTTAT GGCTGCCAGA ACAGCTACAC TCTGAACAAG CTGCTCAAGG 1350
 CCGAGCTGGG CTTCCAGGGC TTTGTCATGA GTGATTGGGC TGCTCACCAT 1400
 GCTGGTGTGA GTGGTGCTTT GGCAGGATTG GATATGTCTA TGCCAGGAGA 1450
 CGTCGACTAC GACAGTGGTA CGTCTTACTG GGGTACAAAC TTGACCATTA 1500
 GCGTGCTCAA CGGAACGGTG CCCCAATGGC GTGTTGATGA CATGGCTGTC 1550
 CGCATCATGG CCGCTACTA CAAGGTCGGC CGTGACCGTC TGTGGACTCC 1600
 TCCCAACTTC AGTCTATGGA CCAGAGATGA ATACGGCTAC AAGTACTACT 1650
 ACCGTCTCGG AGGACCGTAC GAGAAGGTCA ACCAGTACGT GAATGTGCAA 1700
 CGCAACCACA GCGAACTGAT TCGCCGCATT GGAGCGGACA GCACGGTGCT 1750
 CCTCAAGAAC GACGCGGCTC TGCTTTTGAC TGGTAAGGAG CGCCTGGTCG 1800
 CGCTTATCGG AGAAGATGCG GGCTCCAACC CTTATGGTGC CAACGGCTGC 1850
 AGTGACCGTG GATGCGACAA TGGACATTG GCGATGGGCT GGGGAAGTGG 1900
 TACTGCCAAC TTCCCATAC TGGTGACCCC CGAGCAGGCC ATCTCAAACG 1950
 AGGTGCTTAA GCACAAGAA GTGTATTCA CCGCCACCGA TAACTGGGCT 2000
 ATCGATCAAA TTGAGGCGCT TGCTAAGACC GCCAGGTAAG AAGATCCCCG 2050
 ATTCTTTTCC TTCTTGTCGA ATGGATGCTG ACAACATGCT AGTGTCTCTC 2100
 TTGTCTTTGT CAACGCCGAC TCTGGTGAGG GTTACATCAA TGTGGACGGA 2150
 AACCTGGGTG ACCGCAGGAA CCTGACCCTG TGGAGGAACC GCGATAATGT 2200
 GATCAAGGCT GCTGCTAGCA ACTGCAACAA CACAATCGTT GTCATTCACT 2250

4

CTGTCCGACC AGTCTTGTT AACGAGTGGT ACGACAACCC CAATGTTACC 2300
 GCTATCCTCT GGGGTGGTTT GCCCGGTGAG GAGTCTGGCA ACTCTCTTGC 2350
 CGACGTCCCT TATGGCCGTG TCAACCCGGG TGCCAAGTCG CCCTTTTACCT 2400
 GGGGCAAGAC TCGTGAGGCC TACCAAGACT ACTTGGTCAC CGAGCCCAAC 2450
 AACGGCAACG GAGCCCCTCA GGAAGACTTT GTCGAGGGCG TCTTCATTGA 2500
 CTACCGTGGG TTTGACAAGC GCAACGAGAC CCCGATCTAC GAGTTCGGCT 2550
 ATGGTCTGAG CTACACCACT TTCAACTACT CGAACCTTGA GGTGCAGGTG 2600
 CTGAGCGCCC CTGCATACGA GCCTGCTTCG GGTGAGACCG AGGCAGCGCC 2650
 AACCTTCGGA GAGGTTGGAA ATGCGTCGGA TTACCTCTAC CCCAGCGGAT 2700
 TGCTGAGAA TACCAAGTTC ATCTACCCCT GGCTCAACGG TACCGATCTC 2750
 GAGGCATCTT CCGGGGATGC TAGCTACGGG CAGGACTCCT CCGACTATCT 2800
 TCCCGAGGGA GCCACCGATG GCTCTGCGCA ACCGATCCTG CCTGCCGGTG 2850
 GCGGTCTTGG CCGCAACCCT CGCCTGTACG ACGAGCTCAT CCGCGTGTCA 2900
 GTGACCATCA AGAACACCGG CAAGGTGCTT GGTGATGAAG TTCCCCAACT 2950
 GGTAAGTAAA CATGAGGTCC GAACGAGGTT GAACAAAGCT AATCAGTCGC 3000
 AGTATGTTTC CCTTGGCCGT CCAATGAGC CCAAGATCGT GCTGCGTCAA 3050
 TTCGAGCGCA TCACGCTGCA GCCGTCGGAG GAGACGAAGT GGAGCACGAC 3100
 TCTGACGCGC CGTGACCTTG CAAACTGGAA TGTGAGAAG CAGGACTGGG 3150
 AGATTACGTC GTATCCCAAG ATGGTGTGTT TCGGAAGCTC CTCGCGGAAG 3200
 CTGCCGCTCC GGGCGTCTCT GCCTACTGTT CACTAAATAG CTCTCAAATG 3250
 GTATACCATG ATGGCCGTGG TATATGAATT AATGATTAT GCCAACAGCA 3300
 AGACCACTGT AGATGTAGAT GTAGAATGAG TATTGCGTAG TAGCGTGTAG 3350
 ATGATGATAC AAGCGATCCG ACACATGGTA GGAAGAGTGG CGCTAGTTGG 3400
 GGCGGAAACC AAGCGACGTC ATCCGCTGCC GACTTCGCCA GTCTTTCTTC 3450
 TTTTCTCTT CAGCCTTCTT CCTCCGCTTA ATCCAGCAAC CATTGCCAAT 3500
 TGCTCTCTAC ACAAATAAT GCCATAATAC TCTACTCCTA TTCAATATAT 3550
 ACACCACAAT CTCGACATAA TCACACAAGC CTGAACACAC GAGCAACCAT 3600
 GCCCTCTCCC GATCCTCCAG CCCCAGCGAT ACGACCCTTC CAACCACCCA 3650
 TAACAGCGCT CCTCATCTAC CCAGCGACCC TAATCGTGGG ATCACTCTTC 3700
 TCCGTCCTCT CTCACACGCG ACAAGGCACA CGCGACGACG GCTCCAGCAC 3750
 CCTCACCCA CACGTGAGC CCCTAGCCCC GTCCATCGCG TCAGACCTCA 3800
 ACCTCTCCTT TCCTCCGCCG CGCCCCGTCA ACTACTTCGC TCGCAAAGAC 3850
 AACATCTTCA ATCTATATTC GTCAAAGTCG GCTGG 3885

- (2) INFORMATION FOR SEQ ID NO:4:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 6
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:
 Ser Pro Pro Tyr Tyr Pro
 5

- (2) INFORMATION FOR SEQ ID NO:5:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 16
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:
 SNCCNCCNTA YTAYCC 16

- (2) INFORMATION FOR SEQ ID NO:6:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
 Gln Pro Ile Leu Pro Ala Gly Gly
 5

- (2) INFORMATION FOR SEQ ID NO:7:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
 TCCIGCNGGN ARDATNGGYT G 21

- (2) INFORMATION FOR SEQ ID NO:8:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 34
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
 AAACCATGGC TGATGAATTG GCATACTCCC CACC 34

(2) INFORMATION FOR SEQ ID NO:9:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
AAAGGATCCT TAGTGAACAG TAGGCAGAGA CGC 33

(2) INFORMATION FOR SEQ ID NO:10:
(i) SEQUENCE CHARA:
(A) LENGTH: 19
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:
Asp Glu Leu Ala Tyr Ser Pro Pro Tyr Tyr Pro Ser Pro Trp Ala
5 10 15
Asn Gly Gln Gly Asp
20

(2) INFORMATION FOR SEQ ID NO:11:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:
Val Leu Lys His Lys Asn Gly Val Phe Thr Ala Thr Asp Asn Trp
5 10 15
Ala Ile Asp Gln Ile Glu Ala Leu Ala Lys
20 25

(2) INFORMATION FOR SEQ ID NO:12:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:
Gly Ala Thr Asp Gly Ser Ala Gln Pro Ile Leu Pro Ala Gly Gly
5 10 15
Gly Pro Gly Gly Asn Pro
20

(2) INFORMATION FOR SEQ ID NO:13:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3212
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:
GAATTCCCGA TCCTATCTGT CACTTCATCA AAAGGACAGT AGAAAAGGAA 50
GGTGGCACTA CAAATGCCAT CATTGCGATA AAGGAAAGGC TATCGTTCAA 100
GATGCCCTCTG CCGACAGTGG TCCCAAAGAT GGACCCCCAC CCACGAGGAG 150
CATCGTGGA AAGAAGACG TTCCAACCAC GTCTTCAAAG CAAGTGGATT 200
GATGTGATAT CTCCACTGAC GTAAGGGATG ACGCACAAATC CCACTATCCT 250
TCGCAAGACC CTCCTCTAT ATAAGGAAGT TCATTTTCATT TGGAGAGGAC 300
AGGCTTCTTG AGATCCTTCA ACAATTACCA ACAACAACAA ACAACAAACA 350
ACATTACAAT TACTATTTAC AATTACAGTC GACCATGGCT GATGAATTGG 400
CCTACTCCCC ACCGTATTAC CCATCCCCTT GGGCCAATGG CCAGGGCGAC 450
TGGGCGCAGG CATACCAGCG CGCTGTTGAT ATGTCTCGC AAATGACATT 500
GGATGAGAAG GTCAATCTGA CCACAGGAAC TGGATGGGAA TTGGAATAT 550
GTGTTGGTCA GACTGGCGGT GTCCCCGAT TGGGAGTTCC GGAATGTGT 600
TTACAGGATA GCCCTCTGGG CGTTCGCGAC TCCGACTACA ACTCTGCTTT 650
CCCTGCCGGC ATGAACGTGG CTGCGACCTG GGACAAGAAT CTGGCATACC 700

TTCGCGGCAA GGCTATGGGT CAGGAATTTA GTGACAAGGG TGCCGATATC 750
CAATTGGGTC CAGCTGCCGG CCCTCTCGGT AGAAGTCCCG ACGGTGGTCG 800
TAACTGGGAG GGCTTCTCCC CAGACCCTGC CCTAAGTGGT GTGCTCTTTG 850
CCGAGACCAT CAAGGGTATC CAAGATGCTG GTGTGGTTGC GACGGCTAAG 900
CACTACATTG CTTACGAGCA AGAGCATTTT CGTCAGGCGC CTGAAGCCCA 950
AGGTTTTGGA TTTAATATTT CCGAGAGTGG AAGTGCGAAC CTCGATGATA 1000
AGACTATGCA CGAGCTGTAC CTCTGGCCCT TCGCGGATGC CATCCGTGCA 1050
GGTGCTGGCG CTGTGATGTG CTCCTACAAC CAGATCAACA ACAGTTATGG 1100
CTGCCAGAAC AGCTACACTC TGAACAAGCT GCTCAAGGCC GAGCTGGGCT 1150
TCCAGGGCTT TGTCATGAGT GATTGGGCTG CTCACCATGC TGGTGTGAGT 1200
GGTGCTTTGG CAGGATTGGA TATGTCTATG CCAGGAGACG TCGACTACGA 1250
CAGTGGTACG TCTTACTGGG GTACAAACTT GACCATTAGC GTGCTCAACG 1300
GAACGGTGCC CCAATGGCGT GTTGATGACA TGGCTGTCCG CATCATGGCC 1350
GCCTACTACA AGGTCGGCCG TGACCGTCTG TGGACTCCTC CCAACTTCAG 1400
CTCATGGACC AGAGATGAAT ACGGCTACAA GTACTACTAC GTGTCGGAGG 1450
GACCGTACGA GAAGGTCAAC CAGTACGTGA ATGTGCAACG CAACCACAGC 1500
GAACTGATTC GCCGCATTGG AGCGGACAGC ACGGTGCTCC TCAAGAACGA 1550
CGGCGCTCTG CCTTTGACTG GTAAGGAGCG CCTGGTCGCG CTTATCGGAG 1600
AAGATGCGGG CTCCAACCCT TATGGTGCCA ACGGCTGCAG TGACCGTGGA 1650
TGCGACAATG GAACATTGGC GATGGGCTGG GGAAGTGGA CTGCCAACTT 1700
CCCATACCTG GTGACCCCCG AGCAGGCCAT CTCAAACGAG GTGCTTAAGC 1750
ACAAGAATGG TGTATTCACC GCCACCGATA ACTGGGCTAT CGATCAAATT 1800
GAGGCGCTTG CTAAGACCGC CAGTGTCTCT CTTGTCTTTG TCAACGCCGA 1850
CTCTGGTGAG GGTTACATCA ATGTGGACGG AAACCTGGGT GACCGCAGGA 1900
ACCTGACCCT GTGGAGGAAC GGCATAATG TGATCAAGGC TGCTGCTAGC 1950
AACTGCAACA ACACAATCGT TGTCAATCAC TCTGTGGAC CAGTCTTGGT 2000
TAACGAGTGG TACGACAACC CCAATGTTAC CGCTATCCTC TGGGGTGGTT 2050
TGCCCCGTCA GGAGTCTGGC AACTCTCTTG CCGACGTCCT CTATGGCCGT 2100
GTCAACCCCG GTGCCAAGTC GCCCTTTACC TGGGGCAAGA CTCGTGAGGC 2150
CTACCAAGAC TACTTGGTCA CCGAGCCCAA CAACGGCAAC GGAGCCCCTC 2200
AGGAAGACTT TGTCGAGGGC GTCTTCATTG ACTACCGTGG ATTTGACAAG 2250
CGCAACGAGA CCCCATCTA CGAGTTCGGC TATGGTCTGA GCTACACCAC 2300
TTTCAACTAC TCGAACCTTG AGGTGCAGGT GCTGAGCGCC CCTGCATACG 2350
AGCCTGCTTC GGGTGAGACC GAGGCAGCGC CAACCTTCGG AGAGGTTGGA 2400
AATGCGTCGG ATTACCTCTA CCCCAGCGGA TTGCTGAGAA TTACCAAGTT 2450
CATCTACCCC TGGCTCAACG GTACCGATCT CGAGGCATCT TCCGGGGATG 2500
CTAGCTACGG GCAGGACTCC TCCGACTATC TTCCCGAGGG AGCCACCGAT 2550
GGCTCTGCGC AACCATCCT GCCTGCCGGT GCGGTCTCTG GCGGCAACCC 2600
TCGCCTGTAC GACGAGCTCA TCCGCGTGTC AGTGACCATC AAGAACACCG 2650
GCAAGGTTGC TGGTGATGAA GTTCCCCAAC TGTATGTTTC CCTTGGCGGT 2700

7

CCCAATGAGC CCAAGATCGT GCTGCGTCAA TTCGAGCGCA TCACGCTGCA 2750
 GCCGTCGGAG GAGACGAAGT GGAGCAGGAC TCTGACGCGC CGTGACCTTG 2800
 CAAACTGGAA TGTTGAGAAG CAGGACTGGG AGATTACGTC GTATCCCAAG 2850
 ATGGTGTTTG TCGGAAGCTC CTCGCGGAAG CTGCCGCTCC GGGCGTCTCT 2900
 GCCTACTGTT CACTAACCCG GGCGAGCTCG AATTGATCGT TCAAACATTT 2950
 GGCAATAAAG TTTCTTAAGA TTGAATCCTG TTGCCGGTCT TCGGATGATT 3000
 ATCATATAAT TTCTGTTGAA TTACGTTAAG CATGTAATAA TTAAACATGT 3050
 AATGCATGAC GTTATTTATG AGATGGGGTT TTTATGATTA AGAGTCCCCG 3100
 CAATTATACA TTTTAATACG CGATAGAAAA ACAAATATA GCGCCCAAAC 3150
 TAAGGATAAA ATTATTCGCG CCGCGGGGGG GCATTCTATG GTTACTAGAT 3200
 CTCTAGAATT CC 3212

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 841
 (B) TYPE: amino acid
 (C) STRAN: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Asp Glu Leu Ala Tyr Ser Pro Pro Tyr Tyr Pro Ser Pro Trp Ala
 5 10 15
 Asn Gly Gln Gly Asp Trp Ala Gln Ala Tyr Gln Arg Ala Val Asp
 20 25 30
 Ile Val Ser Gln Met Thr Leu Asp Glu Lys Val Asn Leu Thr Thr
 35 40 45
 Gly Thr Gly Trp Glu Leu Glu Leu Cys Val Gly Gln Thr Gly Gly
 50 55 60
 Val Pro Arg Leu Gly Val Pro Gly Met Cys Leu Gln Asp Ser Pro
 65 70 75
 Leu Gly Val Arg Asp Ser Asp Tyr Asn Ser Ala Phe Pro Ala Gly
 80 85 90
 Met Asn Val Ala Ala Thr Trp Asp Lys Asn Leu Ala Tyr Leu Arg
 95 100 105
 Gly Lys Ala Met Gly Gln Glu Phe Ser Asp Lys Gly Ala Asp Ile
 110 115 120
 Gln Leu Gly Pro Ala Ala Gly Pro Leu Gly Arg Ser Pro Asp Gly
 125 130 135
 Gly Arg Asn Trp Glu Gly Phe Ser Pro Asp Pro Ala Leu Ser Gly
 140 145 150
 Val Leu Phe Ala Glu Thr Ile Lys Gly Ile Gln Asp Ala Gly Val
 155 160 165
 Val Ala Thr Ala Lys His Tyr Ile Ala Tyr Glu Gln Glu His Phe
 170 175 180
 Arg Gln Ala Pro Glu Ala Gln Gly Phe Gly Phe Asn Ile Ser Glu
 185 190 195
 Ser Gly Ser Ala Asn Leu Asp Asp Lys Thr Met His Glu Leu Tyr
 200 205 210
 Leu Trp Pro Phe Ala Asp Ala Ile Arg Ala Gly Ala Gly Ala Val
 215 220 225
 Met Cys Ser Tyr Asn Gln Ile Asn Asn Ser Tyr Gly Cys Gln Asn
 230 235 240
 Ser Tyr Thr Leu Asn Lys Leu Leu Lys Ala Glu Leu Gly Phe Gln
 245 250 255
 Gly Phe Val Met Ser Asp Trp Ala Ala His His Ala Gly Val Ser
 260 265 270
 Gly Ala Leu Ala Gly Leu Asp Met Ser Met Pro Gly Asp Val Asp
 275 280 285
 Tyr Asp Ser Gly Thr Ser Tyr Trp Gly Thr Asn Leu Thr Ile Ser
 290 295 300
 Val Leu Asn Gly Thr Val Pro Gln Trp Arg Val Asp Asp Met Ala
 305 310 315
 Val Arg Ile Met Ala Ala Tyr Tyr Lys Val Gly Arg Asp Arg Leu
 320 325 330
 Trp Thr Pro Pro Asn Phe Ser Ser Trp Thr Arg Asp Glu Tyr Gly
 335 340 345
 Tyr Lys Tyr Tyr Tyr Val Ser Glu Gly Pro Tyr Glu Lys Val Asn
 350 355 360
 Gln Tyr Val Asn Val Gln Arg Asn His Ser Glu Leu Ile Arg Arg

8

Ile Gly Ala Asp	365	Ser Thr Val Leu Leu	370	Lys Asn Asp Gly Ala	375
Pro Leu Thr Gly	380	Lys Glu Arg Leu Val	385	Ala Leu Ile Gly Glu	390
Ala Gly Ser Asn	395	Pro Tyr Gly Ala Asn	400	Gly Cys Ser Asp Arg	405
Cys Asp Asn Gly	410	Thr Leu Ala Met Gly	415	Trp Gly Ser Gly Thr	420
Asn Phe Pro Tyr	425	Leu Val Thr Pro Glu	430	Gln Ala Ile Ser Asn	435
Val Leu Lys His	440	Lys Asn Gly Val Phe	445	Thr Ala Thr Asp Asn	450
Ala Ile Asp Gln	455	Ile Glu Ala Leu Ala	460	Lys Thr Ala Ser Val	465
Leu Val Phe Val	470	Asn Ala Asp Ser Gly	475	Glu Gly Tyr Ile Asn	480
Asp Gly Asn Leu	485	Gly Asp Arg Arg Asn	490	Leu Thr Leu Trp Arg	495
Gly Asp Asn Val	500	Ile Lys Ala Ala Ala	505	Ser Asn Cys Asn Asn	510
Ile Val Val Ile	515	His Ser Val Gly Pro	520	Val Leu Val Asn Glu	525
Tyr Asp Asn Pro	530	Asn Val Thr Ala Ile	535	Leu Trp Gly Gly Leu	540
Gly Gln Glu Ser	545	Gly Asn Ser Leu Ala	550	Asp Val Leu Tyr Gly	555
Val Asn Pro Gly	560	Ala Lys Ser Pro Phe	565	Thr Trp Gly Lys Thr	570
Glu Ala Tyr Gln	575	Asp Tyr Leu Val Thr	580	Glu Pro Asn Asn Gly	585
Gly Ala Pro Gln	590	Glu Asp Phe Val Glu	595	Gly Val Phe Ile Asp	600
Arg Gly Phe Asp	605	Lys Arg Asn Glu Thr	610	Pro Ile Tyr Glu Phe	615
Tyr Gly Leu Ser	620	Tyr Thr Thr Phe Asn	625	Tyr Ser Asn Leu Glu	630
Gln Val Leu Ser	635	Ala Pro Ala Tyr Glu	640	Pro Ala Ser Gly Glu	645
Glu Ala Ala Pro	650	Thr Phe Gly Glu Val	655	Gly Asn Ala Ser Asp	660
Leu Tyr Pro Ser	665	Gly Leu Leu Arg Ile	670	Thr Lys Phe Ile Tyr	675
Trp Leu Asn Gly	680	Thr Asp Leu Glu Ala	685	Ser Ser Gly Asp Ala	690
Tyr Gly Gln Asp	695	Ser Ser Asp Tyr Leu	700	Pro Glu Gly Ala Thr	705
Gly Ser Ala Gln	710	Pro Ile Leu Pro Ala	715	Gly Gly Gly Pro Gly	720
Asn Pro Arg Leu	725	Tyr Asp Glu Leu Ile	730	Arg Val Ser Val Thr	735
Lys Asn Thr Gly	740	Lys Val Ala Gly Asp	745	Glu Val Pro Gln Leu	750
Val Ser Leu Gly	755	Gly Pro Asn Glu Pro	760	Lys Ile Val Leu Arg	765
Phe Glu Arg Ile	770	Thr Leu Gln Pro Ser	775	Glu Glu Thr Lys Trp	780
Thr Thr Leu Thr	785	Arg Arg Asp Leu Ala	790	Asn Trp Asn Val Glu	795
Gln Asp Trp Glu	800	Ile Thr Ser Tyr Pro	805	Lys Met Val Phe Val	810
Ser Ser Ser Arg	815	Lys Leu Pro Leu Arg	820	Ala Ser Leu Pro Thr	825
His	830		835		840

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3329
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GAATTCCCGA TCCTATCTGT CACTTCATCA AAAGGACAGT AGAAAAGGAA 50

GGTGGCACTA CAAATGCCAT CATTGCGATA AAGGAAAGGC TATCGTTCAA 100

GATGCCTCTG CCGACAGTGG TCCCAAAGAT GGACCCCCAC CCACGAGGAG 150
CATCGTGGA AAAGAAGACG TTCCAACCAC GTCTTCAAAG CAAGTGGATT 200
GATGTGATAT CTCCACTGAC GTAAGGGATG ACGCACAATC CCACTATCCT 250
TCGCAAGACC CTTCTCTAT ATAAGGAAGT TCATTTCATT TGGAGAGGAC 300
AGGCTTCTTG AGATCCTTCA ACAATTACCA ACAACAACAA ACAACAAACA 350
ACATTACAAT TACTATTTAC AATTACAGTC GAGGGGATCT ATGGCGCGAA 400
AATCCCTAAT TTTCCCGGTG ATTTTGCTCG CCGTTCTTCT CTTCTCTCCG 450
CCGATTTACT CCGCCCGTCA CGATTACCGC GACGCTCTCC GTAAATCTAG 500
CATGGCTGAT GAATTGGCCT ACTCCCCACC GTATTACCCA TCCCCTTGGG 550
CCAATGGCCA GGGCGACTGG GCGCAGGCAT ACCAGCGCGC TGTGATATT 600
GTCTCGCAAA TGACATTGGA TGAGAAGGTC AATCTGACCA CAGGAAGTGG 650
ATGGGAATTG GAACATATGTG TTGGTCAGAC TGGCGGTGTT CCCCATTGG 700
GAGTTCCGGG AATGTGTTTA CAGGATAGCC CTCTGGGCGT TCGCGACTCC 750
GACTACAAC CTGCTTTCCC TGCCGGCATG AACGTGGCTG CGACCTGGGA 800
CAAGAATCTG GCATACCTTC GCGGCAAGGC TATGGGTCAG GAATTTAGTG 850
ACAAGGGTGC CGATATCCAA TTGGGTCCAG CTGCCGGCCC TCTCGGTAGA 900
AGTCCCGACG GTGGTCGTAA CTGGGAGGGC TTCTCCCCAG ACCCTGCCCT 950
AAGTGGTGTG CTCTTTGCCG AGACCATCAA GGGTATCCAA GATGCTGGTG 1000
TGTTTGGCAG GGCTAAGCAC TACATTGCTT ACGAGCAAGA GCATTTCCGT 1050
CAGGCGCCTG AAGCCCAAGG TTTTGGATTT AATATTTCCG AGAGTGAAG 1100
TGCAACCTC GATGATAAGA CTATGCACGA GCTGTACCTC TGGCCCTTCG 1150
CGGATGCCAT CCGTGCAGGT GCTGGCGCTG TGATGTGCTC CTACAACCAG 1200
ATCAACAACA GTTATGGCTG CCAGAACAGC TACACTCTGA ACAAGCTGCT 1250
CAAGGCCGAG CTGGGCTTCC AGGGCTTTGT CATGAGTGAT TGGGCTGCTC 1300
ACCATGCTGG TGTGAGTGGT GCTTTGGCAG GATTGGATAT GTCTATGCCA 1350
GGAGACGTCG ACTACGACAG TGGTACGTCT TACTGGGGTA CAACTTGAC 1400
CATTAGCGTG CTCAACGGAA CGGTGCCCCA ATGGCGTGTT GATGACATGG 1450
CTGTCCGCAT CATGGCCGCC TACTACAAGG TCGCCCGTGA CCGTCTGTGG 1500
ACTCTCCCA ACTTCAGTC ATGGACCAGA GATGAATACG GCTACAAGTA 1550
CTACTACGTG TCGGAGGGAC CGTACGAGAA GGTCAACCAG TACGTGAATG 1600
TGCAACGCAA CCACAGCGAA CTGATTCGCC GCATTGGAGC GGACAGCAGC 1650
GTGCTCCTCA AGAACGACGG CGCTCTGCCT TTGACTGGTA AGGAGCGCCT 1700
GGTCGCGCTT ATCGGAGAAG ATGCGGGCTC CAACCTTAT GGTGCCAACG 1750
GCTGCAGTGA CCGTGGATGC GACAATGGAA CATTGGCGAT GGGCTGGGGA 1800
AGTGGTACTG CCAACTTCCC ATACCTGGTG ACCCCCGAGC AGGCCATCTC 1850
AAACGAGGTG CTTAAGCACA AGAATGGTGT ATTCACCGCC ACCGATAACT 1900
GGGCTATCGA TCAAATTGAG GCGCTTGCTA AGACCGCCAG TGTCTCTCTT 1950
GTCTTTGTCA ACGCCGACTC TGGTGAGGGT TACATCAATG TGGACGGAAA 2000
CCTGGGTGAC CGCAGGAACC TGACCCTGTG GAGGAACGGC GATAATGTGA 2050
TCAAGGCTGC TGCTAGCAAC TGCAACAACA CAATCGTTGT CATTCACTCT 2100

10

GTCGGACCA G TCTTGGTTAA CGAGTGGTAC GACAACCCCA ATGTTACCGC 2150
 TATCCTCTGG GGTGGTTTGC CCGGTCAGGA GTCTGGCAAC TCTCTTGCCG 2200
 ACGTCCTCTA TGGCCGTGTC AACCCCGGTG CCAAGTCGCC CTTTACCTGG 2250
 GGCAAGACTC GTGAGGCCA CCAAGACTAC TTGGTCACCG AGCCCAACAA 2300
 CGGCAACGGA GCCCCTCAGG AAGACTTTGT CGAGGGCGTC TTCATTGACT 2350
 ACCGTGGATT TGACAAGCGC AACGAGACCC CGATCTACGA GTTCGGCTAT 2400
 GGTCTGAGCT ACACCACTTT CAACTACTCG AACCTTGAGG TGCAGGTGCT 2450
 GAGCGCCCT GCATACGAGC CTGCTTCGGG TGAGACCGAG GCAGCGCCAA 2500
 CCTTCGGAGA GGTGGAAAT GCGTCGGATT ACCTCTACCC CAGCGGATTG 2550
 CTGAGAATTA CCAAGTTCAT CTACCCCTGG CTCAACGGTA CCGATCTCGA 2600
 GGCATCTTCC GGGGATGCTA GCTACGGGCA GGA CTCTCTCC GACTATCTTC 2650
 CCGAGGGAGC CACCGATGGC TCTGCGCAAC CGATCCTGCC TGCCGGTGGC 2700
 GGTCTGCGG GCAACCCCTCG CCTGTACGAC GAGCTCATCC GCGTGTCACT 2750
 GACCATCAAG AACACCGGCA AGGTTGCTGG TGATGAAGTT CCCCAACTGT 2800
 ATGTTTCCCT TGGCGGTCCC AATGAGCCCA AGATCGTGCT GCGTCAATTC 2850
 GAGCGCATCA CGCTGCAGCC GTCGGAGGAG ACGAAGTGGA GCACGACTCT 2900
 GACGCGCCGT GACCTTGCAA ACTGGAATGT TGAGAAGCAG GACTGGGAGA 2950
 TTACGTCGTA TCCCAAGATG GTGTTTGTGCG GAAGCTCCTC GCGGAAGCTG 3000
 CCGCTCCGGG CGTCTCTGCC TACTGTTCAC TAACCCGGGC GAGCTCGAAT 3050
 TGATCGTTCA AACATTGGC AATAAAGTTT CTTAAGATTG AATCCTGTTG 3100
 CCGGTCTTGC GATGATTATC ATATAATTTC TGTGAATTA CGTTAAGCAT 3150
 GTAATAATTA AACATGTAAT GCATGACGTT ATTTATGAGA TGGGGTTTTT 3200
 ATGATTAAGA GTCCCCGCAA TTATACATTT TAATACGCGA TAGAAAAACA 3250
 AAATATAGCG CCCAACTAA GGATAAAATT ATTCGCGCCG CGGGGGGGCA 3300
 TTCTATGGTT ACTAGATCTC TAGAATTCC 3329

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 880
 (B) TYPE: amino acid
 (C) STRAN: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met Ala Arg Lys Ser Leu Ile Phe Pro Val Ile Leu Leu Ala Val
 5 10 15
 Leu leu Phe Ser Pro Pro Ile Tyr Ser Ala Gly His Asp Tyr Arg
 20 25 30
 Asp Ala Leu Arg Lys Ser Ser Met Ala Asp Glu Leu Ala Tyr Ser
 35 40 45
 Pro Pro Tyr Tyr Pro Ser Pro Trp Ala Asn Gly Gln Gly Asp Trp
 50 55 60
 Ala Gln Ala Tyr Gln Arg Ala Val Asp Ile Val Ser Gln Met Thr
 65 70 75
 Leu Asp Glu Lys Val Asn Leu Thr Thr Gly Thr Gly Trp Glu Leu
 80 85 90
 Glu Leu Cys Val Gly Gln Thr Gly Gly Val Pro Arg Leu Gly Val
 95 100 105
 Pro Gly Met Cys Leu Gln Asp Ser Pro Leu Gly Val Arg Asp Ser
 110 115 120
 Asp Tyr Asn Ser Ala Phe Pro Ala Gly Met Asn Val Ala Ala Thr
 125 130 135
 Trp Asp Lys Asn Leu Ala Tyr Leu Arg Gly Lys Ala Met Gly Gln
 140 145 150
 Glu Phe Ser Asp Lys Gly Ala Asp Ile Gln Leu Gly Pro Ala Ala

	155		160		165
Gly Pro Leu Gly	Arg Ser Pro Asp Gly	Gly Arg Asn Trp Glu	Gly		
	170		175		180
Phe Ser Pro Asp	Pro Ala Leu Ser Gly	Val Leu Phe Ala Glu	Thr		
	185		190		195
Ile Lys Gly Ile	Gln Asp Ala Gly Val	Val Ala Thr Ala Lys	His		
	200		205		210
Tyr Ile Ala Tyr	Glu Gln Glu His Phe	Arg Gln Ala Pro Glu	Ala		
	215		220		225
Gln Gly Phe Gly	Phe Asn Ile Ser Glu	Ser Gly Ser Ala Asn	Leu		
	230		235		240
Asp Asp Lys Thr	Met His Glu Leu Tyr	Leu Trp Pro Phe Ala	Asp		
	245		250		255
Ala Ile Arg Ala	Gly Ala Gly Ala Val	Met Cys Ser Tyr Asn	Gln		
	260		265		270
Ile Asn Asn Ser	Tyr Gly Cys Gln Asn	Ser Tyr Thr Leu Asn	Lys		
	275		280		285
Leu Leu Lys Ala	Glu Leu Gly Phe Gln	Gly Phe Val Met Ser	Asp		
	290		295		300
Trp Ala Ala His	His Ala Gly Val Ser	Gly Ala Leu Ala Gly	Leu		
	305		310		315
Asp Met Ser Met	Pro Gly Asp Val Asp	Tyr Asp Ser Gly Thr	Ser		
	320		325		330
Tyr Trp Gly Thr	Asn Leu Thr Ile Ser	Val Leu Asn Gly Thr	Val		
	335		340		345
Pro Gln Trp Arg	Val Asp Asp Met Ala	Val Arg Ile Met Ala	Ala		
	350		355		360
Tyr Tyr Lys Val	Gly Arg Asp Arg Leu	Trp Thr Pro Pro Asn	Phe		
	365		370		375
Ser Ser Trp Thr	Arg Asp Glu Tyr Gly	Tyr Lys Tyr Tyr Tyr	Val		
	380		385		390
Ser Glu Gly Pro	Tyr Glu Lys Val Asn	Gln Tyr Val Asn Val	Gln		
	395		400		405
Arg Asn His Ser	Glu Leu Ile Arg Arg	Ile Gly Ala Asp Ser	Thr		
	410		415		420
Val Leu Leu Lys	Asn Asp Gly Ala Leu	Pro Leu Thr Gly Lys	Glu		
	425		430		435
Arg Leu Val Ala	Leu Ile Gly Glu Asp	Ala Gly Ser Asn Pro	Tyr		
	440		445		450
Gly Ala Asn Gly	Cys Ser Asp Arg Gly	Cys Asp Asn Gly Thr	Leu		
	455		460		465
Ala Met Gly Trp	Gly Ser Gly Thr Ala	Asn Phe Pro Tyr Leu	Val		
	470		475		480
Thr Pro Glu Gln	Ala Ile Ser Asn Glu	Val Leu Lys His Lys	Asn		
	485		490		495
Gly Val Phe Thr	Ala Thr Asp Asn Trp	Ala Ile Asp Gln Ile	Glu		
	500		505		510
Ala Leu Ala Lys	Thr Ala Ser Val Ser	Leu Val Phe Val Asn	Ala		
	515		520		525
Asp Ser Gly Glu	Gly Tyr Ile Asn Val	Asp Gly Asn Leu Gly	Asp		
	530		535		540
Arg Arg Asn Leu	Thr Leu Trp Arg Asn	Gly Asp Asn Val Ile	Lys		
	545		550		555
Ala Ala Ala Ser	Asn Cys Asn Asn Thr	Ile Val Val Ile His	Ser		
	560		565		570
Val Gly Pro Val	Leu Val Asn Glu Trp	Tyr Asp Asn Pro Asn	Val		
	575		580		585
Thr Ala Ile Leu	Trp Gly Gly Leu Pro	Gly Gln Glu Ser Gly	Asn		
	590		595		600
Ser Leu Ala Asp	Val Leu Tyr Gly Arg	Val Asn Pro Gly Ala	Lys		
	605		610		615
Ser Pro Phe Thr	Trp Gly Lys Thr Arg	Glu Ala Tyr Gln Asp	Tyr		
	620		625		630
Leu Val Thr Glu	Pro Asn Asn Gly Asn	Gly Ala Pro Gln Glu	Asp		
	635		640		645
Phe Val Glu Gly	Val Phe Ile Asp Tyr	Arg Gly Phe Asp Lys	Arg		
	650		655		660
Asn Glu Thr Pro	Ile Tyr Glu Phe Gly	Tyr Gly Leu Ser Tyr	Thr		
	665		670		675
Thr Phe Asn Tyr	Ser Asn Leu Glu Val	Gln Val Leu Ser Ala	Pro		
	680		685		690
Ala Tyr Glu Pro	Ala Ser Gly Glu Thr	Glu Ala Ala Pro Thr	Phe		
	695		700		705
Gly Glu Val Gly	Asn Ala Ser Asp Tyr	Leu Tyr Pro Ser Gly	Leu		
	710		715		720
Leu Arg Ile Thr	Lys Phe Ile Tyr Pro	Trp Leu Asn Gly Thr	Asp		
	725		730		735
Leu Glu Ala Ser	Ser Gly Asp Ala Ser	Tyr Gly Gln Asp Ser	Ser		
	740		745		750
Asp Tyr Leu Pro	Glu Gly Ala Thr Asp	Gly Ser Ala Gln Pro	Ile		

12

755	760	765
Leu Pro Ala Gly	Gly Gly Pro Gly Gly Asn Pro Arg Leu Tyr	Asp
770	775	780
Glu Leu Ile Arg	Val Ser Val Thr Ile Lys Asn Thr Gly Lys	Val
785	790	795
Ala Gly Asp Glu	Val Pro Gln Leu Tyr Val Ser Leu Gly Gly	Pro
800	805	810
Asn Glu Pro Lys	Ile Val Leu Arg Gln Phe Glu Arg Ile Thr	Leu
815	820	825
Gln Pro Ser Glu	Glu Thr Lys Trp Ser Thr Thr Leu Thr Arg	Arg
830	835	840
Asp Leu Ala Asn	Trp Asn Val Glu Lys Gln Asp Trp Glu Ile	Thr
845	850	855
Ser Tyr Pro Lys	Met Val Phe Val Gly Ser Ser Ser Arg Lys	Leu
860	865	870
Pro Leu Arg Ala	Ser Leu Pro Thr Val His	
875	880	

(2) INFORMATION FOR SEQ ID NO:17

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

His Asp Glu Leu

(2) INFORMATION FOR SEQ ID NO:18

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3338
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GAATTCCTCGA TCCTATCTGT CACTTCATCA AAAGGACAGT AGAAAAGGAA 50
 GGTGGCACTA CAAATGCCAT CATTGCGATA AAGGAAAGGC TATCGTTCAA 100
 GATGCCTCTG CCGACAGTGG TCCCAAAGAT GGACCCCCAC CCACGAGGAG 150
 CATCGTGGAA AAAGAAGACG TTCCAACCAC GTCTTCAAAG CAAGTGGATT 200
 GATGTGATAT CTCCACTGAC GTAAGGGATG ACGCACAATC CCACTATCCT 250
 TCGCAAGACC CTTCCTCTAT ATAAGGAAGT TCATTTTATT TGGAGAGGAC 300
 AGGCTTCTTG AGATCCTTCA ACAATTACCA ACAACAACAA ACAACAAACA 350
 ACATTACAAT TACTATTTAC AATTACAGTC GAGGGGATCT ATGGCGCGAA 400
 AATCCCTAAT TTTCCCGGTG ATTTTGCTCG CCGTTCTTCT CTTCTCTCCG 450
 CCGATTTACT CCGCCGGTCA CGATTACCGC GACGCTCTCC GTAAATCTAG 500
 CATGGCTGAT GAATTGGCCT ACTCCCCACC GTATTACCCA TCCCCTTGGG 550
 CCAATGGCCA GGGCGACTGG GCGCAGGCAT ACCAGCGCGC TGTGATATT 600
 GTCTCGCAAA TGACATTGGA TGAGAAGGTC AATCTGACCA CAGGAACTGG 650
 ATGGGAATTG GAACTATGTG TTGGTCAGAC TGGCGGTGTT CCCCATTGG 700
 GAGTTCCGGG AATGTGTTTA CAGGATAGCC CTCTGGGCGT TCGCGACTCC 750
 GACTACAACCT CTGCTTTCCC TGCCGGCATG AACGTGGCTG CGACCTGGGA 800
 CAAGAATCTG GCATACCTTC GCGGCAAGGC TATGGGTCAG GAATTTAGTG 850
 ACAAGGGTGC CGATATCCAA TTGGGTCCAG CTGCCGGCCC TCTCGGTAGA 900
 AGTCCCGACG GTGTCGTAA CTGGGAGGGC TTCTCCCCAG ACCCTGCCCT 950
 AAGTGGTGTG CTCTTTGCGG AGACCATCAA GGGTATCCAA GATGCTGGTG 1000
 TGGTTGCGAC GGCTAAGCAC TACATTGCTT ACGAGCAAGA GCATTTCCGT 1050
 CAGCGCCTG AAGCCCAAGG TTTTGGATTT AATATTTCCG AGAGTGGAAG 1100

TGCGAACCTC GATGATAAGA CTATGCACGA GCTGTACCTC TGGCCCTTCG 1150
CGGATGCCAT CCGTGCAGGT GCTGGCGCTG TGATGTGCTC CTACAACCAG 1200
ATCAACAACA GTTATGGCTG CCAGAACAGC TACTCTCTGA ACAAGCTGCT 1250
CAAGGCCGAG CTGGGCTTCC AGGGCTTTGT CATGAGTGAT TGGGCTGCTC 1300
ACCATGCTGG TGTGAGTGGT GCTTTGGCAG GATTGGATAT GTCTATGCCA 1350
GGAGACGTCG ACTACGACAG TGGTACGTCT TACTGGGGTA CAAACTTGAC 1400
CATTAGCGTG CTCAACGGAA CGGTGCCCCA ATGGCGTGTT GATGACATGG 1450
CTGTCCGCAT CATGGCCGCC TACTACAAGG TCGGCCGTGA CCGTCTGTGG 1500
ACTCTCCCA ACTTCAGCTC ATGGACCAGA GATGAATACG GCTACAAGTA 1550
CTACTACGTG TCGGAGGGAC CGTACGAGAA GGTCAACCAG TACGTGAATG 1600
TGCAACGCAA CCACAGCGAA CTGATTCGCC GCATTGGAGC GGACAGCACG 1650
GTGCTCCTCA AGAACGACGG CGCTCTGCCT TTGACTGGTA AGGAGCGCCT 1700
GGTCGCGCTT ATCGGAGAAG ATGCGGGCTC CAACCCTTAT GGTGCCAACG 1750
GCTGCAGTGA CCGTGGATGC GACAATGGAA CATTGGCGAT GGGCTGGGGA 1800
AGTGGTACTG CCAACTTCCC ATACCTGGTG ACCCCCGAGC AGGCCATCTC 1850
AAACGAGGTG CTTAAGCACA AGAATGGTGT ATTCACCGCC ACCGATAACT 1900
GGGCTATCGA TCAAATTGAG GCGCTTGCTA AGACCGCCAG TGTCTCTCTT 1950
GTCTTTGTCA ACGCCGACTC TGGTGAGGGT TACATCAATG TGGACGGAAA 2000
CCTGGGTGAC CGCAGGAACC TGACCCTGTG GAGGAACGGC GATAATGTGA 2050
TCAAGGCTGC TGCTAGCAAC TGCAACAACA CAATCGTTGT CATTCACTCT 2100
GTGGGACCAG TCTTGTTAA CGAGTGGTAC GACAACCCCA ATGTTACCGC 2150
TATCCTCTGG GGTGGTTTGC CCGGTCAGGA GTCTGGCAAC TCTCTTGCCG 2200
ACGTCTCTA TGGCCGTGTC AACCCCGGTG CCAAGTCGCC CTTTACCTGG 2250
GGCAAGACTC GTGAGGCCTA CCAAGACTAC TTGGTCACCG AGCCCAACAA 2300
CGGCAACGGA GCCCTCAGG AAGACTTTGT CGAGGGCGTC TTCATTGACT 2350
ACCGTGGATT TGACAAGCGC AACGAGACCC CGATCTACGA GTTCGGCTAT 2400
GGTCTGAGCT ACACCACTTT CAACTACTCG AACCTTGAGG TGCAGGTGCT 2450
GAGCGCCCCT GCATACGAGC CTGCTTCGGG TGAGACCGAG GCAGCGCCAA 2500
CCTTCGGAGA GGTGGAAAT GCGTCGGATT ACCTCTACCC CAGCGGATTG 2550
CTGAGAATTA CCAAGTTCAT CTACCCCTGG CTCAACGGTA CCGATCTCGA 2600
GGCATCTTCC GGGGATGCTA GCTACGGGCA GGACTCCTCC GACTATCTTC 2650
CCGAGGGAGC CACCGATGGC TCTGCGCAAC CGATCCTGCC TGCCGGTGGC 2700
GGTCCTGGCG GCAACCCCTG CCTGTACGAC GAGCTCATCC GCGTGTCACT 2750
GACCATCAAG AACACCGGCA AGGTTGCTGG TGATGAAGTT CCCCAACTGT 2800
ATGTTTCCCT TGGCGGTCCC AATGAGCCCA AGATCGTGCT GCGTCAATTC 2850
GAGCGCATCA CGCTGCAGCC GTCGGAGGAG ACGAAGTGGA GCACGACTCT 2900
GACGCGCCGT GACCTTGCAA ACTGGAATGT TGAGAAGCAG GACTGGGAGA 2950
TTACGTCGTA TCCCAAGATG GTGTTTGTG GAAGCTCCTC GCGGAAGCTG 3000
CCGCTCCGGG CGTCTCTGCC TACTGTTCAT GATGAACTTT AACCCGGGCG 3050
AGCTCGAATT GATCGTTCAA ACATTGGCA ATAAAGTTTC TTAAGATTGA 3100

GTTAAGCATG TAATAATTAA ACATGTAATG CATGACGTTA TTTATGAGAT 3200
 GGGGTTTTTA TGATTAAGAG TCCCCGCAAT TATACATTTT AATACGCGAT 3250
 AGAAAAACAA AATATAGCGC CCAAATAAG GATAAAATTA TTCGCGCCGC 3300
 GGGGGGGCAT TCTATGGTTA CTAGATCTCT AGAATTCC

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 883
 (B) : amino acid
 (C) STRAN: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Met Ala Arg Lys Ser Leu Ile Phe Pro Val Ile Leu Leu Ala Val
 5 10 15
 Leu leu Phe Ser Pro Pro Ile Tyr Ser Ala Gly His Asp Tyr Arg
 20 25 30
 Asp Ala Leu Arg Lys Ser Ser Met Ala Asp Glu Leu Ala Tyr Ser
 35 40 45
 Pro Pro Tyr Tyr Pro Ser Pro Trp Ala Asn Gly Gln Gly Asp Trp
 50 55 60
 Ala Gln Ala Tyr Gln Arg Ala Val Asp Ile Val Ser Gln Met Thr
 65 70 75
 Leu Asp Glu Lys Val Asn Leu Thr Thr Gly Thr Gly Trp Glu Leu
 80 85 90
 Glu Leu Cys Val Gly Gln Thr Gly Gly Val Pro Arg Leu Gly Val
 95 100 105
 Pro Gly Met Cys Leu Gln Asp Ser Pro Leu Gly Val Arg Asp Ser
 110 115 120
 Asp Tyr Asn Ser Ala Phe Pro Ala Gly Met Asn Val Ala Ala Thr
 125 130 135
 Trp Asp Lys Asn Leu Ala Tyr Leu Arg Gly Lys Ala Met Gly Gln
 140 145 150
 Glu Phe Ser Asp Lys Gly Ala Asp Ile Gln Leu Gly Pro Ala Ala
 155 160 165
 Gly Pro Leu Gly Arg Ser Pro Asp Gly Gly Arg Asn Trp Glu Gly
 170 175 180
 Phe Ser Pro Asp Pro Ala Leu Ser Gly Val Leu Phe Ala Glu Thr
 185 190 195
 Ile Lys Gly Ile Gln Asp Ala Gly Val Val Ala Thr Ala Lys His
 200 205 210
 Tyr Ile Ala Tyr Glu Gln Glu His Phe Arg Gln Ala Pro Glu Ala
 215 220 225
 Gln Gly Phe Gly Phe Asn Ile Ser Glu Ser Gly Ser Ala Asn Leu
 230 235 240
 Asp Asp Lys Thr Met His Glu Leu Tyr Leu Trp Pro Phe Ala Asp
 245 250 255
 Ala Ile Arg Ala Gly Ala Gly Ala Val Met Cys Ser Tyr Asn Gln
 260 265 270
 Ile Asn Asn Ser Tyr Gly Cys Gln Asn Ser Tyr Thr Leu Asn Lys
 275 280 285
 Leu Leu Lys Ala Glu Leu Gly Phe Gln Gly Phe Val Met Ser Asp
 290 295 300
 Trp Ala Ala His His Ala Gly Val Ser Gly Ala Leu Ala Gly Leu
 305 310 315
 Asp Met Ser Met Pro Gly Asp Val Asp Tyr Asp Ser Gly Thr Ser
 320 325 330
 Tyr Trp Gly Thr Asn Leu Thr Ile Ser Val Leu Asn Gly Thr Val
 335 340 345
 Pro Gln Trp Arg Val Asp Asp Met Ala Val Arg Ile Met Ala Ala
 350 355 360
 Tyr Tyr Lys Val Gly Arg Asp Arg Leu Trp Thr Pro Pro Asn Phe
 365 370 375
 Ser Ser Trp Thr Arg Asp Glu Tyr Gly Tyr Lys Tyr Tyr Tyr Val
 380 385 390
 Ser Glu Gly Pro Tyr Glu Lys Val Asn Gln Tyr Val Asn Val Gln
 395 400 405
 Arg Asn His Ser Glu Leu Ile Arg Arg Ile Gly Ala Asp Ser Thr
 410 415 420
 Val Leu Leu Lys Asn Asp Gly Ala Leu Pro Leu Thr Gly Lys Glu
 425 430 435
 Arg Leu Val Ala Leu Ile Gly Glu Asp Ala Gly Ser Asn Pro Tyr
 440 445 450
 Gly Ala Asn Gly Cys Ser Asp Arg Gly Cys Asp Asn Gly Thr Leu
 455 460 465

15

Ala Met Gly Trp	Gly Ser Gly Thr Ala	Asn Phe Pro Tyr Leu	Val
470	475	480	
Thr Pro Glu Gln	Ala Ile Ser Asn Glu	Val Leu Lys His Lys	Asn
485	490	495	
Gly Val Phe Thr	Ala Thr Asp Asn Trp	Ala Ile Asp Gln Ile	Glu
500	505	510	
Ala Leu Ala Lys	Thr Ala Ser Val Ser	Leu Val Phe Val Asn	Ala
515	520	525	
Asp Ser Gly Glu	Gly Tyr Ile Asn Val	Asp Gly Asn Leu Gly	Asp
530	535	540	
Arg Arg Asn Leu	Thr Leu Trp Arg Asn	Gly Asp Asn Val Ile	Lys
545	550	555	
Ala Ala Ala Ser	Asn Cys Asn Asn Thr	Ile Val Val Ile His	Ser
560	565	570	
Val Gly Pro Val	Leu Val Asn Glu Trp	Tyr Asp Asn Pro Asn	Val
575	580	585	
Thr Ala Ile Leu	Trp Gly Gly Leu Pro	Gly Gln Glu Ser Gly	Asn
590	595	600	
Ser Leu Ala Asp	Val Leu Tyr Gly Arg	Val Asn Pro Gly Ala	Lys
605	610	615	
Ser Pro Phe Thr	Trp Gly Lys Thr Arg	Glu Ala Tyr Gln Asp	Tyr
620	625	630	
Leu Val Thr Glu	Pro Asn Asn Gly Asn	Gly Ala Pro Gln Glu	Asp
635	640	645	
Phe Val Glu Gly	Val Phe Ile Asp Tyr	Arg Gly Phe Asp Lys	Arg
650	655	660	
Asn Glu Thr Pro	Ile Tyr Glu Phe Gly	Tyr Gly Leu Ser Tyr	Thr
665	670	675	
Thr Phe Asn Tyr	Ser Asn Leu Glu Val	Gln Val Leu Ser Ala	Pro
680	685	690	
Ala Tyr Glu Pro	Ala Ser Gly Glu Thr	Glu Ala Ala Pro Thr	Phe
695	700	705	
Gly Glu Val Gly	Asn Ala Ser Asp Tyr	Leu Tyr Pro Ser Gly	Leu
710	715	720	
Leu Arg Ile Thr	Lys Phe Ile Tyr Pro	Trp Leu Asn Gly Thr	Asp
725	730	735	
Leu Glu Ala Ser	Ser Gly Asp Ala Ser	Tyr Gly Gln Asp Ser	Ser
740	745	750	
Asp Tyr Leu Pro	Glu Gly Ala Thr Asp	Gly Ser Ala Gln Pro	Ile
755	760	765	
Leu Pro Ala Gly	Gly Gly Pro Gly Gly	Asn Pro Arg Leu Tyr	Asp
770	775	780	
Glu Leu Ile Arg	Val Ser Val Thr Ile	Lys Asn Thr Gly Lys	Val
785	790	795	
Ala Gly Asp Glu	Val Pro Gln Leu Tyr	Val Ser Leu Gly Gly	Pro
800	805	810	
Asn Glu Pro Lys	Ile Val Leu Arg Gln	Phe Glu Arg Ile Thr	Leu
815	820	825	
Gln Pro Ser Glu	Glu Thr Lys Trp Ser	Thr Thr Leu Thr Arg	Arg
830	835	840	
Asp Leu Ala Asn	Trp Asn Val Glu Lys	Gln Asp Trp Glu Ile	Thr
845	850	855	
Ser Tyr Pro Lys	Met Val Phe Val Gly	Ser Ser Ser Arg Lys	Leu
860	865	870	
Pro Leu Arg Ala	Ser Leu Pro Thr Val	His Asp Glu Leu	
875	880		

(2) INFORMATION FOR SEQ ID NO:20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

CAGTGACCGT GGATGCGACA ATG 23

(2) INFORMATION FOR SEQ ID NO:21

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AGAGACGGAT GACAAGTACT ACTTGAAATT GGGCCCAAAA 40

(2) INFORMATION FOR SEQ ID NO:22

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single

16

(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:
CAGTGACCGT GGATGCGACA ATG 23

(2) INFORMATION FOR SEQ ID NO:23

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:
AAAGGATCCT TAGTGAACAG TAGGCAGAGA CGC 33

(2) INFORMATION FOR SEQ ID NO:24

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:
His Asp Glu Leu

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/36586 A3

- (51) International Patent Classification⁷: C12N 5/04, 3 Lipkin Street, 76411 Rehovot (IL). SHOSEYOV, Oded (IL/IL); 5 Haerez Street, 72910 Karme Yossef (IL).
15/09, 15/31, 15/82, A01H 5/00
- (21) International Application Number: PCT/IL00/00758 (74) Agent: G. E. EHRLICH (1995) LTD.; 28 Bezalel Street, 52521 Ramat Gan (IL).
- (22) International Filing Date: 15 November 2000 (15.11.2000) (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 09/443,338 19 November 1999 (19.11.1999) US (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant (for all designated States except US): YISSUM RESEARCH DEVELOPMENT COMPANY OF THE HEBREW UNIVERSITY OF JERUSALEM [IL/IL]; P.O. Box 4279, 46 Jabotinsky Street, 91042 Jerusalem (IL).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): SHU, Wei [CN/IL]; 88/5 Etzel Road, 76361 Rehovot (IL). MARTON, Ira [IL/IL]; 10 Aharonovich Yoseph Street, 76347 Rehovot (IL). SIEGEL, Daniel, L. [IL/IL]; 11/5 Weizmann Street, 76280 Rehovot (IL). BEN-AMI, Bravdo [IL/IL]; 11 Han-kin Street, 76354 Rehovot (IL). DEKEL, Mara [IL/IL];
- Published:
— with international search report
- (88) Date of publication of the international search report: 2 May 2002
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: ASPEGILLUS NIGER BETA-GLUCOSIDASE GENE, PROTEIN AND USES THEREOF

(57) Abstract: A polypeptide having β -glucosidase enzymatic activity, a polynucleotide encoding the polypeptide, a nucleic acid constructs carrying the polynucleotide, transformed or infected cells, such as yeast cells, and transgenic organisms expressing the polynucleotide and various uses of the polypeptide, the polynucleotide, cells and/or organisms, including, producing a recombinant polypeptide having the β -glucosidase enzymatic activity, increasing the level of aroma compounds in alcoholic beverages, as well as other fermentation products of plant material, hydrolyzing cellobiose and thus increasing the level of fermentable glucose, increasing the production of alcohol, such as ethanol from plant material, increasing the aroma released from a plant or a plant product, and hydrolysis or transglycosylation of glycosides.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL00/00758

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 5/04; 15/09, 15/31, 15/82; A01H 5/00

US CL : Please See Extra Sheet.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 800/278, 288, 295, 298; 536/23.2, 23.7; 435/69.1, 320.1, 419, 468

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WEST 2.0, STN CAS

search terms: beta-glucosidase, Aspergillus niger, E. coli, fermentation, aroma, free-glucose,

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	PENTTILA et al. Cloning of Aspergillus niger Genes in Yeast. Expression of the Gene Coding Aspergillus Beta-glucosidase. Mol. Gen. Genet. 1984, Vol. 194, pages 494-499, see pages 494-497.	1, 3-10, 13-17, 20-23, 43, 49, 61-63
Y	US 5,997,913 A (FOWLER et al) 07 December 1999 (07.12.99), see entire document, especially columns 4-5, 23-30	1, 3-11, 13-17, 20-23, 26, 29, 32-33, 36-37, 40-43, 46, 49, 52-55, 58, 61-63, 66-107

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*G* document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means	
P document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

30 MAY 2001

Date of mailing of the international search report

21 JUN 2001

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

MEDINA A. IBRAHIM

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL00/00758

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to cla
Y,P	US 6,015,703 A (WHITE et al) 18 January 2000 (18.01.00), see entire document.	1, 3-11, 13-1 20-23, 26, 29 33, 36-37, 40 46, 49, 52-55 61-63, 66-10
Y	EP 0307 071 A2 (YISSUM RESEARCH DEVELOPMENT COMPANY of the HEBREW UNIVERSITY of JERUSALEM) 15 March 1989 (15.03.89), see entire document.	66-107

Form PCT/ISA/210 (continuation of second sheet) (July 1998) *

INTERNATIONAL SEARCH REPORT

International application No.
PCT/IL00/00758**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: 2, 12, 18, 19, 24-25, 27-28, 31, 34-35, 38-39, 44-45, 47-48, 50-51, 56-57, 59-60, 64-65,
because they relate to parts of the international application that do not comply with the prescribed requirements to such
an extent that no meaningful international search can be carried out, specifically:
- The claims recite SEQ ID NO:s and no CRF has been furnished. Therefore, no meaningful search report could be
established.
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable
claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment
of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers
only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is
restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐
☐

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet(1)) (July 1998) *

INTERNATIONAL SEARCH REPORT

International application No
PCT/IL00/00758

A. CLASSIFICATION OF SUBJECT MATTER:

US CL :

800/278, 288, 295, 298; 536/23.2, 23.7; 435/69.1, 320.1, 419, 468

Form PCT/ISA/210 (extra sheet) (July 1998) *

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☒ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.